THE UNIVERSITY
of EDINBURGH

Investigation of Mechanisms of DNA Methylation Maintenance Regulation in Pluripotency

David Hay

THE UNIVERSITY
of EDINBURGH

Degree of Doctor of Philosophy

The University of Edinburgh

2019

Declaration

I declare that this thesis has been composed by me, and that all of the work is my own unless otherwise stated.

David Hay

September 2019

Acknowledgements

Firstly, thank you to Maria for giving me the opportunity to study for a PhD and be a member of her lab. It has been an enjoyable, challenging and rewarding experience all at the same time, during which I have learned and developed a lot. Thanks for your guidance throughout.

Thank you to all the MAC Lab members past and present – Christine, Denitsa, John, Tom, Vinay – for all the suggestions and support. It has been great to be a member of a wonderful team.

Also, it may sound cliched but there is no better way to say it, Abby, I really could not have done it without you – thanks so much!

Thank you to Richard Meehan for his constant support and for sharing his knowledge on all things DNA methylation related and beyond. The Meehan lab members, past and present, have all been extremely generous with their time and have helped with reagents, protocols, data analysis and advice – thanks to all of them.

Thank you to Ian Adams and Joe Marsh for their scrutiny on my thesis committee panel which helped shape the direction of the project. The feedback and advice given by PIs in section and joint lab meetings was also very much appreciated.

Throughout my PhD I have used a variety of methods and techniques, so thank you to Andy Finch, Alex von Kriegsheim and Jimi Wills for their help with mass spec and Harris Morrison with cell imaging. Thank you also to Noor Gammoh for her time in showing me how to analyse proteomic data.

Experiments and day-to-day working were facilitated greatly by the technical services and support staff. Similarly, thank you to everyone in the Genome Regulation section for helping to troubleshoot my experiments and lending reagents.

Thank you to all the mentors and teachers I have had that made learning and the prospect of discovering new things exciting.

Finally, thank you to my friends and family for putting up with me for the past four years and your invaluable words of encouragement that kept me going.

Abstract

The early stages of development are characterised by profound transcriptional and epigenetic reprogramming. The most striking example is the global loss of DNA methylation patterns on the maternal and paternal genomes that is precipitated upon formation of the zygote. This is coincident with the establishment of the inner-cell mass, a pluripotent population of cells that give rise to the embryo. The necessity to reset the epigenetic landscape to attain this state however requires further investigation, as are the mechanisms that lead to these changes. Using a mouse embryonic stem culture media, referred to as 2i, that models this developmental period, I set out to uncover mechanisms that may contribute to the remodelling of DNA methylation. In contrast to traditional serum-based mouse embryonic stem cell culture media, 2i maintains DNA methylation at a significantly lower level. The use of 2i and serum cultures therefore provided a means to investigate the regulation of the different DNA methylation landscapes. The overarching focus has been to understand how the mediators of DNA methylation, the DNA methyltransferases, are regulated at the level of the protein in response to the stimuli of the media. The differential makeup of the protein complexes of Dnmt1, referred to as the main DNA methyltransferase, and of its obligate co-factor, Uhrf1, were characterised in a proteomic screen. This led to the identification of several candidates and their potential roles in mediating DNA methylation have been postulated to build a case for future investigation. Furthermore, the influence of citrullination, a relatively understudied post-translational modification detected throughout early development, was also investigated in these model culture systems.

Lay summary

DNA can be viewed as a sequence of the letters A, T, G and C, bundled into one of the smallest units the human body can be divided into, a cell. These letters can be arranged into specific sequences termed genes. In turn, genes are read by machinery in a cell to produce proteins, the fundamental building blocks of life. A full complement of these letters is initially provided equally by sperm and egg cells from biological males and females, respectively. These two cells each have half of the DNA required for a new organism to develop. When they come together by a process known as fertilisation, a full set of genes is present. Having a full set of genes is in itself not sufficient and the precise reading of these at certain times or in different types of cells is also crucial for life. One way this is controlled in adults is through the use of small chemical marks that can directly decorate DNA. These can dictate if and when a gene produces a protein in any certain cell. One example of these marks is DNA methylation. This is present on the male and female DNA in sperm and eggs, however it is nearly completely removed after fertilisation, before it is added back again later. The necessity for this initial erasure and the way it happens is still not clear. This study has set out to identify the mechanisms that promote this loss by focussing on the machinery that regulates DNA methylation levels. This machinery may be modified directly or controlled by other machinery in the cell to keep regulating DNA methylation. When these mechanisms have been identified, they can be promoted or prevented from happening before their effects on the development of life can be concluded.

Contents

Abbreviations

5-aza-CdR	5-aza-2’-deoxycytidine
5caC	5-carboxylcytosine
5fC	5-formyluracil
5hmC	5-hydroxymethylcytosine
5hmU	5-hydroxymethyluracil
5mC	5-methylcytosine
5'UTR	5’ untranslated region
ADD	Alpha thalassaemia/mental retardation syndrome X-linked homologue (ATRX)–DNMT3–DNMT3L
AdoMet	S-adenosyl-L-methionine
AID	Activation-induced deaminase
AML	Acute myeloid leukaemia
AMP	Adenosine monophosphate
AMPK	AMP-activated protein kinase
APE1	AP endonuclease 1
Apobec3	Apolipoprotein B editing complex 3
BAH	Bromo adjacent homology
BER	Base excision repair
bp	Base pair
BSA	Bovine serum albumin
CAG	CMV early enhancer/chicken β actin
CAP	CXXC affinity purification
CDCA7	Cell division cycle associated 7
CGI	CpG Island
CTCF	CCCTC-binding factor
DAPI	4’,6-diamidino-2-phenylindole
DMAP1	Dnmt1-associated protein 1
DMEM	Dulbecco's Modified Eagle Medium
DMR	Differentially methylated region

Dnmt	DNA-(cytosine-5) methyltransferase
Dnmt1o	Oocyte Dnmt1
Dnmt1s	Somatic Dnmt1
DNMTi	DNMT inhibitors
E2	17β-estradiol
EpiLC	Epiblast-like cell
EpiLSC	Epiblast stem cell
EpiSCs	Epiblast stem cells
ERV	Endogenous retroviruses
Ethanol	EtOH
FCS	Fetal calf serum
FGF4	Fibroblast growth factor 4
FRAP	Fluorescence Recovery After Photobleaching
GK	Glycine-Lysine
Glasgow's MEM	GMEM
Gsk3β	Glycogen synthase kinase 3β
H3	Histone 3
H3K9me	Histone 3 lysine 9 methylation
H3R	Histone 3 arginine
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HIV-1	Human immunodeficiency virus type 1
HP1- β	Heterochromatin protein 1- β
HRP	Horseradish peroxidase
HUVEC	Human umbilical vein endothelial cell
IAP	Intracisternal A-particle
ICF	Immunodeficiency, centromere instability and facial anomalies
ICM	Inner-cell mass
ICR	Imprint control region
Id	Inhibitor of Differentiation

IP	Immunoprecipitation
iPSCs	Induced-pluripotent cells
ITC	Isothermal titration calorimetry
KAP1	KRAB-associated protein 1
KOSR	KnockOut™ Serum Replacement
KRAB-ZFP	Kruppel-associated box-containing ZFP
LC-MS	Liquid-chromatography mass spectrometry
LIG1	DNA ligase 1
LincRNAs	Long intergenic non-coding RNAs
LINE-1	Long interspersed element class 1
lncRNA	Long non-coding RNA
Lsh	Lymphocyte-specific helicase
MBP	Methyl-CpG binding protein
MEF	Mouse embryonic fibroblast
Mek	Mitogen-activation protein kinase
MEL	Murine erythroleukemia
mESCs	Mouse embryonic stem cells
MIN	Multifunctional integrase
miRNA	Micro-RNA
MS	Multiple sclerosis
MSCV	Murine stem cell virus
NEAAs	Non-essential amino acids
NET	Neutrophil extracellular trap
NRF	Nuclear respiratory factor
OSKM	Oct4, Sox2, Klf4, c-Myc
OSN	Oct4, Sox2, Nanog
PADI	Peptidylarginine deiminase
PARP	Poly ADP-ribose polymerase
PBD	Pcna binding domain
Pcna	Proliferating cell nuclear-antigen

Pdc	Pyruvate dehydrogenase complex
Pdk1	Pyruvate dehydrogenase kinase 1
PFA	Paraformaldehyde
PGC	Primordial germ cell
PGC-1a	Peroxisome proliferator-activated receptor gamma coactivator-1a
PHD	Plant homeobox domain
PKB	Protein kinase B
PolII	Polymerase II
PTM	Post-translational modification
Rb	Retinoblastoma
RFTS	Replication foci-targeting sequence
RING	Really Interesting New Gene
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction
SINE	Short interspersed elements
sncRNA	Small non-coding RNA
SRA	SET and RING associated
SUMO	Small ubiquitin-related modifier
TBS	Tris-buffered saline
TDG	Thymine DNA glycosylase
TET	Ten-eleven-translocation
TKO	Triple (Dnmt) knockout
TopII-α	Topoisomerase II-α
TRD	Target recognition domain
TSG101	Tumour susceptibility gene 101
TSS	Transcriptional start sites
TTD	Tandem tudor domain
UBL	Ubiqutin-like Domain
UbqC	Ubiquitin C
UDG	Uracil DNA glycosylase
Uhrf1	Ubiquitin-like, containing PHD and RING finger domains 1

UNG2	Uracil glycosylase
Usp7	Ubiquitin-specific protease 7
Utf1	Undifferentiated Embryonic Cell Transcription Factor 1
WT	Wild-type
ZGA	Zygotic gene activation
Zfp	Zinc finger protein
Zscan	Zinc finger and SCAN domain-containing protein

Chapter 1 – Introduction

1.1 Background

Methylated cytosine is described in the classic Riggs, Pugh and Holliday papers as an epigenetic factor that regulates and constrains cellular fates (Holliday and Pugh, 1975; Riggs, 1975). Epigenetic programs act in concert with spatiotemporally coordinated interactions between transcription factors and DNA to give rise to specialised cells (Baubec and Schübeler, 2014). Gene regulation is primarily encoded in *cis* whilst DNA methylation is a mitotically heritable modification that does not alter the underling sequence of DNA. These marks are essential in mammalian development and as such, DNA methylation has spurned a field into elucidating its roles (Smith and Meissner, 2013).

Although present in organisms as diverse as plants, animals and prokaryotes, DNA methylation is not an essential requirement for life and is absent from common model organisms, such as adult *Drosophila melanogaster, Caenorhabditis elegans* and *Saccharomyces cerevisiae* (Kumar *et al.*, 1994; Hendrich and Tweedie, 2003).

The mammalian genome is densely methylated at cytosine bases in a palindromic CpG context. This is to such an extent in humans that 70-80% of CpGs are modified, which equates to 1% of total DNA bases (Bird, 2002). Methylation has also been observed at very low levels in a CHG and CHH configuration in oocytes and neurones, where H represents any other nucleotide base, which may represent the fidelity of the enzymes that deposit the mark (Li, Dai, Suzanne N Martos, *et al.*, 2015; C. Li *et al.*, 2018).

The landscape in most invertebrates is markedly different. For example, in the sea squirt *Ciona intestinalis,* methylation is present as blocks interspersed with stretches of unmethylated DNA in equal proportions. Gene bodies of highly expressed genes and transposable elements, described later in more detail, are marked in accordance with the pattern of surrounding DNA. The fungi *Neurospora crassa* exhibits a similar mosaic pattern however methylation is targeted specifically to transposable elements. This is also the case for plants such as *Arabidopsis thaliana* wherein marks are enriched over gene bodies of many housekeeping genes too. On the contrary, plants with relatively larger genomes have a broad global distribution of DNA methylation, potentially owing to the abundance of transposable elements embedded in the genome that are targeted to maintain a transcriptionally repressive state (Suzuki and Bird, 2008).

The shift from a mosaic to a global pattern in mammals has been presented as necessary for the development of an immune system that could detect the DNA of bacterial pathogens, which is replete with unmethylated CpG. The Toll-like receptor 9 expressed by immune cells in vertebrates has an affinity for this motif and is absent in invertebrates. This notion is certainly provocative as it suggests that the majority of DNA methylation is redundant and is present largely in a global configuration to obscure actual functional blocks, such as across gene bodies or transposable elements, that would otherwise trigger an autoimmune response. The implication therefore is that outside of specific targets, DNA methylation has no role (Suzuki and Bird, 2008). However, this would ignore the selectivity of the enzymes that initially set up the DNA methylation patterns and the influence of co-factors and histone marks that can direct them.

As our understanding is enhanced by the development of a plethora of sequencing technologies, the initial importance attributed to DNA methylation has been slowly eroded. Nonetheless, DNA methylation is irrefutably vital for post implantation embryonic development. This is illustrated by the lethality of knockout mouse models for the enzymes which set up and maintain the patterns across cell division (Li, Bestor and Jaenisch, 1992; Okano *et al.*, 1999).

1.2 DNA Methyltransferases

DNA-(cytosine-5) methyltransferases (Dnmts) catalyse and maintain DNA methylation. They are defined by ten conserved motif blocks that are invariably ordered along the protein and punctuated with non-conserved variable sequences that can determine target specificity (see Figure 1.1) (Kumar *et al.*, 1994). These fold into a characteristic 'AdoMet-dependent MTase fold' and constitute the C-terminus catalytic domain (Jeltsch and Jurkowska, 2016).

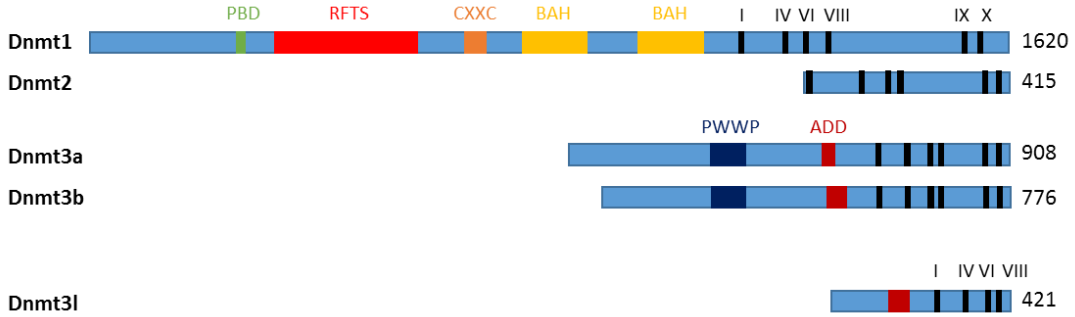


Figure 1.1 – Domains of the mammalian DNA methyltransferases. Present in all are important motifs that constitute the catalytic domain (I, IV, VI, VIII, IX, X). Two of these motifs (IX, X) are notably absent in Dnmt3l, which renders it catalytically inactive. Details of other domains labelled are discussed later in the text. Values to the right of each schematic refer to the number of residues in the mouse methyltransferase. Adapted from (Goll and Bestor, 2005; Jurkowska, Jurkowski and Jeltsch, 2011)

Motif IV PCQ contains a nucleophilic cysteine residue that reacts with the C6 position in a cytosine ring. The nitrogenous base is also transiently protonated at the N3 position, a state stabilised by motif VI ENV of the MTase fold. This results in the formation of a transient covalent protein-DNA complex and the activation of C5 in the cytosine ring, leading to a nucleophilic attack on the donor S-adenosyl-*L*-methionine (AdoMet) methyl group. Subsequent deprotonation at the C5 position, followed by β-elimination, regenerates the aromatic ring and the by-product *S*-adenosyl-*L*-homocysteine (see Figure 1.2) (Kumar *et al.*, 1994; Jurkowska, Jurkowski and Jeltsch, 2011; Jeltsch *et al.*, 2017).

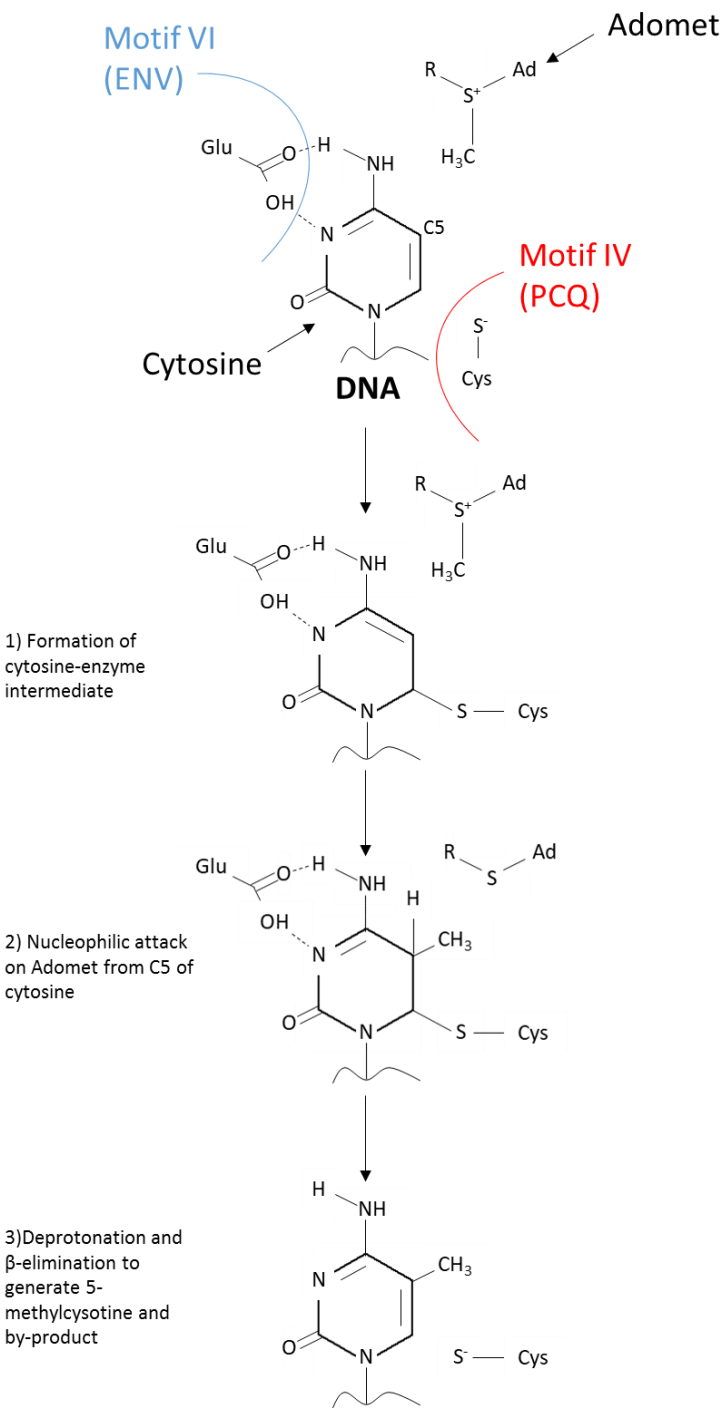


Figure 1.2 – Cytosine methylation is catalysed by DNA methyltransferases in a multi-step process as outlined above. See main body of text for further details (Adapted from Vilkaitis *et al.*, 2001; Jeltsch, 2006)

Mammalian DNA methylation patterns are initially deposited by *de novo* methyltransferases Dnmt3a and -b which act on DNA devoid of pre-existing marks. The maintenance methyltransferase Dnmt1 meanwhile has an affinity for hemi-methylated sites that arise

during cell division. It semi-conservatively copies these patterns onto nascent daughter strands using the parental strand as a template (see Figure 1.3) (Smith and Meissner, 2013).

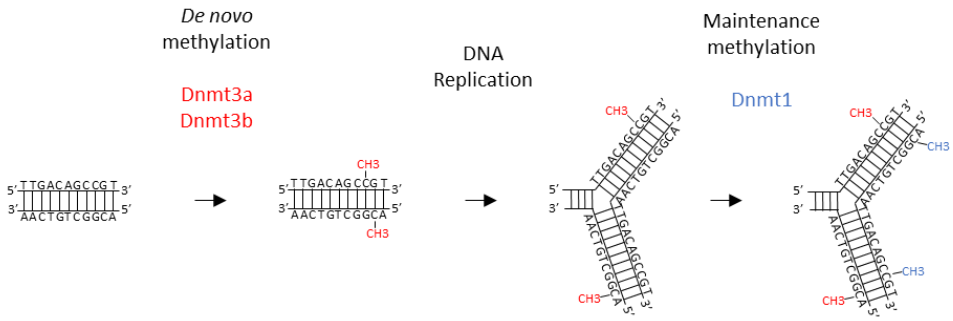


Figure 1.3 – When DNA is initially synthesised, it is devoid of methyl groups. Marks are added through the action of the *de novo* methyltransferases, typically at CpG sequences. During DNA replication, the parental strand is used as a template by Dnmt1 to semi-conservatively copy the methylation marks onto the newly synthesised strand, thus the pattern is maintained in daughter cells. Adapted from: Moore, Le and Fan, 2012.

Underscoring the significance of DNA methylation for development are mouse models. Homozygous *Dnmt1*-null mice die by E10.5 with apparent limb-bud defects, impaired organogenesis, open neural pore, increased cell death and no yolk sac vasculature (Li, Bestor and Jaenisch, 1992). A compound heterozygote null/catalytically dead mouse model has an indistinguishable phenotype and arrests developmentally at the same stage as a null line (Takebayashi *et al.*, 2007). In both instances this has been partially explained by the aberrant expression of imprinted genes and transposable elements which will be discussed later (Walsh and Bestor, 1999; Takebayashi *et al.*, 2007). Depletion of *Dnmt1* in mouse embryonic fibroblasts (MEFs) results in apoptosis within 5-6 days and this can be partially attenuated by the knockdown of *p53* (Jackson-Grusby *et al.*, 2001).

The effects of deleting *de novo* transferases varies: *Dnmt3b*-null mice die by E9.5 which contrasts with *Dnmt3a*-null mice that survive until after birth. These are however runted in size and succumb within four weeks. Recovered *Dnmt3b*-/- embryos are characterised by growth impairment and a variety of developmental defects (Okano *et al.*, 1999). Noteworthy is that heterozygotes from the single knockout of either the maintenance or *de novo*

transferases are phenotypically indistinguishable from wild-type (WT) mice (Li, Bestor and Jaenisch, 1992; M Okano *et al.*, 1999).

Mouse embryonic stem cells (mESCs) exhibit a high expression of Dnmts in addition to components not yet introduced, Dnmt3l and the highly specific Dnmt2 enzyme. Despite this, they are tolerant of epigenetic perturbations and can be cultured in perpetuity in appropriate conditions, as exemplified by *Dnmt* single or triple knockout models (TKOs) (Tsumura *et al.*, 2006; Meissner, 2010). This contrasts with human ESCs from which *DNMT1*-homozygous null models cannot be derived, whilst single and double *DNMT3A* and *-B* clones can. This disparity may be a consequence of the different cell-doubling times which is substantially longer for hESCs. Nevertheless, like the mouse equivalents, early passage *DNMT3A* and *-B* knockouts can form teratomas (Liao *et al.*, 2015). A tolerance to genetic deletions and response to rescues, coupled with an ability for multilineage differentiation, mean mESCs have been used extensively to grapple with questions surrounding the importance of methylation in development and regulation of the Dnmts. Consequently, they are the model of choice throughout this project and are referred to extensively herein.

1.3 Roles of DNA methylation

DNA methylation is highly associated with transcriptional regulation. This function may be direct, through the occlusion of transcription factors, or secondary whereby methylated CpGs act as docking sites for methyl-CpG binding proteins (MBPs). These serve as recruitment factors and bring about a transcriptionally repressive state that inhibits initiation, typically through hypoacetylation of neighbouring histones (Schübeler *et al.*, 2000; Fournier *et al.*, 2012). MBPs are expressed ubiquitously in somatic cells and also have tissue specific isoforms (Wade, 2001).

For example, MeCP2, an X-linked gene from the MBP family, transiently recruits the Sin3-histone deacetylase (HDAC) repressor complex and regulates spacing of linker histones (Fournier *et al.*, 2012). MeCP2 is highly abundant in post-natal neuronal tissue and conditional deletion in mice leads to the development of a phenotype that is comparable to human patients with Rett syndrome, a disorder stemming from mutations in *MeCP2* that abolish its DNA methylation binding capabilities. The phenotype is characterised by post-natal neurological defects (Chen *et al.*, 2001; Guy *et al.*, 2001; Wade, 2001). Alternative

mutations in MeCP2 can disrupt interactions with ATRX, a SWI/SNF DNA helicase that cooperates with cohesin and the transcription factor CCCTC-binding factor (CTCF) to maintain the imprint status of a subset of genes in the mouse brain (Fournier *et al.*, 2012). Despite this, only a small number of genes are miss-expressed by 10-20% in Rett syndrome patients and *MBP* knockout models only produce modest effects (Hendrich and Tweedie, 2003). The association with DNA methylation is complicated by observations that MeCP2 can bind to unmethylated DNA and compact nucleosome arrays *in vitro* and assemble oligomeric suprastructures (Georgel *et al.*, 2003).

Kaiso and Kaiso-like proteins similarly work to repress transcription by altering histone states. Their methyl-dependent binding instigates hypoacetylation through recruitment of CtBP1 and CtBP2 complexes (Sasai and Defossez, 2009).

That methylation inhibits transcription factor binding however has proven to be generalisation and is motif, context and transcription factor dependent. Although it is instructive at highly dense CpG regions with respect to transcriptional activity, as discussed below, the effect of DNA methylation is variable within other promoters (Baubec and Schübeler, 2014).

Using DNAseI hypersensitivity as a surrogate for transcription factor binding, accessible sites were shown to be largely comparable between mESCs in which the three main Dnmts have been knocked-out, TKOs, and the WT counterpart (Domcke *et al.*, 2015). A number of sites did emerge however that were distal to transcriptional start sites (TSS). These had a notably low CpG content and were enriched for motifs of the transcription factors GABPA, MYCN and most prominently, NRF1. The subsequent binding of NRF1 was coincident with increased transcription of the nearest gene. In WT cells, the engagement of NRF1 was shown to be directed by other adjacently binding transcription factors insensitive to methylation, such as REST, that act in *trans* to promote local demethylation of NRF1 motifs (Domcke *et al.*, 2015).

A similarly minimal effect on CTCF redistribution was also found in a colorectal carcinoma cell line (HCT116) null for *Dnmt1* and *Dnmt3b* (DKO) (Maurano *et al.*, 2015). In excess of 98.5% of unoccupied CTCF sites in the WT HCT116 line remained depleted of the transcription factor in the DKO line. Whilst over 3000 sites were found to arise exclusively with demethylation, only 50% of sites methylated in WT lines and depleted in the DKO lines displayed this novel binding. This suggests that methylation contributes but is not the sole determinant of transcription factor binding (Maurano *et al.*, 2015).

Techniques such as methylation-sensitive SELEX and EpiSELEX-seq have begun to unravel the complex relationship between transcription factor binding and DNA methylation (Kribelbauer *et al.*, 2017; Yin *et al.*, 2017).

A systematic analysis of the interplay between DNA methylation and the binding affinities of 542 human transcription factors was performed with methylation-sensitive SELEX (Yin *et al.*, 2017). Over a third of transcription factors were shown to not be influenced by methylation due the absence of a CpG sequence in the primary binding motif, whilst others were not affected regardless of a methyl group. A further third of transcription factors, enriched for examples involved in development, displayed a preference for the methylated motif that likely interacts via hydrophobic interactions to the DNA binding domain. Around a quarter, typically involved in cell proliferation and differentiation, preferred an unmethylated substrate due to steric clashes that would otherwise arise with a methyl-group. As a single motif is recognised by multiple transcription factors that are either positively or negatively influenced by DNA methylation, this property can select the factors that bind and this may impose different outcomes on gene expression (Yin *et al.*, 2017).

The positional effects of methylation must also be considered if more than one CpG dinucleotide in a transcription factor binding motif is present. Transcription factors can also form complexes with co-factors and recruit enzymes that remodel the 5-methylcytosine (5mC) levels and alter binding affinity (Feldmann *et al.*, 2013; Kribelbauer *et al.*, 2017). Furthermore, motifs that lack a CpG site may be indirectly regulated by DNA methylation if MBPs are recruited to a neighbouring methylated sequence and their activity encroaches, inhibiting transcription factor binding (Baubec and Schübeler, 2014).

Examples of genes that are causally regulated by DNA methylation and exhibit no interplay with other epigenetic regulatory mechanisms have been identified (Hackett *et al.*, 2012). Exposure of NIH 3T3 fibroblasts to the demethylating agent 5-aza-2'-deoxycytidine (5-aza-CdR) resulted in a significant over 6-fold upregulation of 344 genes and 49 remained expressed after a period of recovery. These were distinguished as possessing a CpG Island (CGI) promoter, a high density CpG promoter as explained below. Cross referencing this set of 49 genes with transcripts upregulated in Dnmt1 hypomorphic MEFs produced 26 genes that are likely bona fide examples of DNA methylation dependent genes. Half of these were germ-line specific genes, with a number involved in transposable element suppression. Their silencing occurred in parallel with their methylation during ESC differentiation and

conversely, expression of these genes was detected during the wave of DNA methylation loss that occurs *in vivo* during primordial germ cell (PGC) specification (Hackett *et al.*, 2012). That most genes remain silent when DNA methylation is removed shows that regulation is primarily determined by the presence of a transcription factor in the cell or alternative epigenetic modifications that act at the level of histones (O'Neill *et al.*, 2018).

1.3.1 CpG Islands

Exceptions to when CpGs dyads are typically methylated can be found at high density clusters termed CGIs, which are refractory to methylation. Various calculations agree that over two-thirds of mammalian promoters could be defined as such, particularly housekeeping and key developmental genes and lineage specific microRNAs. Consequently, there is considerable interest in them as they are potential elegant illustrations of the inhibitory function of DNA methylation on gene expression (Meissner *et al.*, 2008; Illingworth and Bird, 2009; Lienert *et al.*, 2011). The description of a CGI, and therefore the parameters inputted into genome browsers to quantify them, often varies in terms of the length however, there is consensus that the G + C content is ≥ 50% and an observed/expected CpG ratio of ≥0.6 (Illingworth and Bird, 2009). Compared to low CpG density promoters, CGI associated promoters in mESCs display higher transcriptional activity. This appears to be due to an intrinsic property of CpG-dense sequences that reside in and that flank transcription factor motifs (Hartl *et al.*, 2019).

Why these sequences remain largely unmethylated under a potential onslaught of Dnmt activity is likely due to the presence of bound transcription factors, hence the relationship between house-keeping genes and hypomethylated CGIs. Transcription factor binding may occlude Dnmts or bring about a loss of any DNA methylation that is added by an active process, introduced later (Brandeis *et al.*, 1994; Feldmann *et al.*, 2013; Marchal and Miotto, 2015; Yin *et al.*, 2017). The binding of transcription factors alone was shown to be insufficient to cause transcription and therefore this hypothesis could also account for the unmethylated status of CGIs in other gene classes (Spitz and Furlong, 2012). However, as CGIs largely remain depleted of 5mC regardless of transcriptional activity, this points to additional regulatory mechanisms that enforce the typical hypomethylated state (Brandeis *et al.*, 1994; Weber *et al.*, 2007). Indeed, H3K4me3 enrichment at CGIs, the mechanistic reasons for which are addressed later, occludes binding of the *de novo* methyltransferases (Mikkelsen *et al.*, 2007; Weber *et al.*, 2007; Meissner *et al.*, 2008; Illingworth and Bird, 2009). A component of the Setd1 H3K4 transferase, Cfp1, was identified as having a high affinity for CGIs, thus providing

a link between the underlying DNA and perepuation of a DNA methylation resistant region (Thomson *et al.*, 2010).

Exploiting the affinity of the CXXC zinc-finger domain for unmethylated and clustered CpG sites, CXXC affinity purification (CAP)-seq was developed to identify CGIs in human tissue. The total number found (~25,200) exceeded the number of known genes (~20,000) whilst gene rich chromosomes, such as chromosome 19, had more CGIs compared to gene poor chromosomes, for example chromosome 18. This demonstrated that CGIs are not exclusively localised to canonical promoters (Illingworth *et al.*, 2008). CGIs that lay outside of these units were termed orphan CGIs. Typically, the GC content of murine CGIs is lower than humans but are equally abundant in number and distributed throughout the gene, including at annotated promoters and inter- and intra-genic positions (Illingworth *et al.*, 2010).

Data from the CAP-seq study in brain, blood, muscle and spleen tissue showed that tissue-specific methylation of CGI associated promoters is enriched at genes encoding transcription factors and key developmental genes. In general, the presence of methylation was shown to be a poor predictor of gene expression (Illingworth *et al.*, 2008). In a different study using human primary fibroblasts however, whilst there was no significant correlation with methylation at low CpG content promoters and polymerase II (PolII) binding, indicative of transcription, decreased PolII occupancy was shown at hypermethylated CGIs (Weber *et al.*, 2007). These findings suggest caution when considering CGI methylation as regulating a dichotomy of silencing or expression. Studies into orphan-CGIs have been more resolute and have shown that ~40% of human and mouse units display transcriptional activity and these are more likely to become methylated and silenced during cell specification, compared to those overlapping an annotated promoter (Illingworth *et al.*, 2010).

1.3.2 Imprinting

The influence of DNA methylation at CGIs is less ambiguous in the context of genomic imprinting. Differential methylation at CGI containing imprint control regions (ICRs), giving rise to a differentially methylated regions (DMRs), in conjunction with histone modifications and non-coding RNA binding, ensures expression in a parental-origin-specific manner from either the maternal or paternal chromosome. These DMRs regulate multiple genes in a cluster and result in functional non-equivalence of the genomes, showing maternal and paternal genomes provide unique reciprocal functions. Likewise, perturbing individual

imprinted genes on the chromosomes from which they are expressed leads to developmental defects (Ferguson-Smith, 2011).

Although imprinting has been described as a useful paradigm for exploring interactions between gene expression and epigenetics, the arrangement of imprinted genes at clusters and the influence of neighbouring elements ensures methylation confers differential effects on expression. This may be expected as over 100 imprinted genes have been confirmed so far in the mouse (Ferguson-Smith, 2011). One of the best studied paternally methylated ICRs is that which controls *H19-Igf2* reciprocity. In females, a shared enhancer upstream of the genes engages with *H19* as CTCF insulates *Igf2* situated downstream. Conversely, deposition of methylation blocks CTCF binding in males, which is concomitant with *Igf2* expression (Leighton *et al.*, 1995; Thorvaldsen, Duran and Bartolomei, 1998).

Germline-acquired methylation represents the canonical method for the establishment of imprints. The inability of rescued *Dnmt1 -/-* mESCs to restore imprint status demonstrates this (Tucker *et al.*, 1996). However, tissue specific imprinting can emerge in the adult. For example, at the *Dlk1-Dio3* locus, methylation is present at the ICR on the paternal chromosome in almost all tissues where it drives the expression of *Dlk1*. In neural stem cells, it also found on the maternal chromosome which was experimentally determined to be essential for the self-renewal property in adult male mice (Ferrón *et al.*, 2011).

1.3.3 X-chromosome inactivation

X-chromosome inactivation in eutherian females describes the random silencing of genes on one X-chromosome, thus achieving parity with males which only inherit one X-chromosome. This is initiated by *Xist,* a non-coding RNA expressed from an X-inactivation centre. Coating of the X-chromosome from which it was expressed in *cis* accumulates to an extent that by E8.5, all female somatic cells have this feature. A correlation between methylation at a CpG-rich promoter region upstream of *Xist* and silencing is observed on the active X-chromosome, in contrast to a methylation free domain on the inactive X, demonstrates a regulatory role of DNA methylation (Panning and Jaenisch, 1996). Perversely, the active X-chromosome is relatively hypermethylated. However, this can be accounted for by the prevalence of the mark over gene bodies whereas the inactive X-chromosome is solely enriched at promoters (Weber *et al.*, 2005).

Although initially expressed in both male and female mESCs in culture, *Xist* is extinguished in differentiated male cells. However, cells from a *Dnmt1 -/-* background maintain expression and this temporally correlates with apoptosis (Panning and Jaenisch, 1996). X-chromosome inactivation can nevertheless be initiated normally in the absence of Dnmt3a and -b, with female embryos at E9.5 exhibiting monoallelic expression of *Xist* that is capable of repressing transcription from the chromosome *in cis*. The implication therefore is that DNA methylation does not instigate the differential silencing of *Xist* however, it is required for stable silencing (Sado, 2004). Binding of Mbd2 to 5mC has shown to contribute to the enforcement of a transcriptionally inert state by recruiting the NuRD complex which mediates histone hypoacetylation (Barr *et al.*, 2007).

1.3.4 Repression of DNA repeats

The majority of the mammalian genome is composed of non-coding sequences yet in some instances, it is essential that DNA methylation is still present. Regulating the potential transcriptional activity and stability of repetitive sequences has been regarded as the function for which DNA methylation first evolved (Yoder, Walsh and Bestor, 1997). This notion is strengthened by the apparent conservation of this role in diverse organisms (Walsh, Chaillet and Bestor, 1998; Zemach *et al.*, 2010; Bewick *et al.*, 2016). Like in other instances of transcriptional repression, DNA methylation may impair the expression of repetitive sequences through steric hindrance or recruitment of repressive MBPs (Crichton *et al.*, 2014).

A repetitive sequence can refer to satellite DNA or transposable elements. Satellite DNA is concentrated at centromeres and may possess a structural role, as evinced in ICF syndrome (discussed below). It can be further classified as major or minor satellite DNA, in accordance with length (Du, Lianna M. Johnson, *et al.*, 2015). Transposable elements meanwhile are sequences found in most organisms that have the potential to move independently throughout the genome, possibly through an RNA intermediate. Strikingly, these compromise ~40% of sequenced mammalian genomes (International Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium *et al.*, 2002) and derivations are frequently associated with gene promoters (Jordan *et al.*, 2003).

The burden of regulating a class of these transposable elements, endogenous retroviruses (ERVs), is less in humans in which they are rarely mobile. This contrasts with mice wherein 10% of spontaneous mutations are directly caused by their mobilisation and deleterious insertion into the genome. These events have the potential to cause ectopic activation, as

the element can serve as a promoter, or insertional mutagenesis. One of the best studied elements of this class is the intracisternal A-particle (IAP), up to 2000 copies of which have been estimated to reside in the murine genome (Walsh, Chaillet and Bestor, 1998). In a classic experiment, induced demethylation of an IAP element residing upstream of an *agouti* gene was sufficient to breed a higher proportion of yellow-coated mice, owing to a greater expression of the gene. This demonstrated the capacity of dysregulated IAP elements to modulate gene expression (Gaudet *et al.*, 2004).

Elements that are potentially active in both humans and mice are exemplified by long interspersed element class 1 (LINE-1). Proteins required to promote the activity of these are encoded within the elements and can be subsequently utilised by short interspersed elements (SINEs) that are present throughout mouse and human genomes also (Garcia-Perez, Widmann and Adams, 2016). Whereas SINEs follow the general rule of non-repetitive sequences, irrespective of CpG density LINEs are hypermethylated (Meissner *et al.*, 2008). The correlation between DNA methylation and LINE-1 activity is witnessed in the hypomethylated placental tissue, which derives from the trophectoderm of the developing blastocyst, and neuronal precursor cells that are moderately hypomethylated relative to somatic tissue. Notably, LINE-1 reporters indicate that retrotransposition occurs more frequently in these tissues and activity in neurones may contribute to the mosaicism observed in brain tissue (Garcia-Perez, Widmann and Adams, 2016).

The acute removal of DNA methylation in mESCs can be induced with culture in media referred to as 2i, as discussed later. In one study, the 2i media was further supplemented with vitamin C and the effects on transposable element methylation status and expression were examined (Walter *et al.*, 2016). It was noted that loss of methylation progressed at differing rates for different classes of elements: ERVs were most resistant as 5mC was maintained across the element, whilst there was only a retention at the 5' untranslated region (5'UTR) of LINE-1 elements. Nevertheless, this coincided with a burst in transcription of a proportion of intact elements. In contrast, there is only a modest increase of a subset of ERVs in TKO mESCs (Karimi *et al.*, 2011; Walter *et al.*, 2016). In 2i + vitamin C media, elements became suppressed almost universally at the same time. Exceptions to this trend included IAPs and some SINEs which remained consistently derepressed. Re-silencing was shown to be largely driven by histone 3 lysine 9 trimethylation (H3K9me3) and H3K27me3 accumulation at adjacent genomic sequences and across or at flanking regions of the

transposons. The extent and combination of H3K9me3 and H3K27me3 acquisition were element and family specific. The compensation by suppressive histone marks therefore provides an explanation for the disparity between acute DNA methylation removal in 2i + vitamin C and long-term cultured TKO cells on transposable element expression (Karimi *et al.*, 2011; Walter *et al.*, 2016). Nevertheless, transcriptional repression in *Dnmt3a -/-* ;*Dnmt3b-/-* mESCs harbouring murine stem cell virus (MSCV) and human immunodeficiency virus type 1 (HIV-1) vectors suggests that DNA methylation is a secondary event to silencing and therefore the conditions used in the study above may impose additional effects not accounted for (Pannell *et al.*, 2000).

With respect to IAP, the direct relationship between demethylation and derepression has been resolved with an additional intermediate step. This was initially borne from observing a discrepancy between *Dnmt1*- and *Uhrf1*-null mESCs. Uhrf1 (Ubiquitin-like, containing PHD and RING finger domains 1) is an obligate co-factor of Dnmt1 and is discussed in depth later. Here however, it is important to note that it recruits Dnmt1 to chromatin by binding to hemi-methylated sites that arise during cell division and these would therefore persist in the absence of Dnmt1. In a *Dnmt1*-null mESCs, there is a transient increase in the expression of an IAP subgroup whereas in the corresponding *Uhrf1*-null line, methylation is reduced to levels comparable to the *Dnmt1* mutants and IAP expression is only modestly and transiently promoted. This was accounted for by the maintained binding of Uhrf1 to hemi-methylated sites across IAP that arise in *Dnmt1*-null cells. This subsequently inhibited the recruitment of Setd1, a H3K9me3 transferase, and the silencing of IAP. As cell division progresses, hemimethylated sites are diluted, permitting Setd1 binding. That this only affects a specific subclass of IAP elements was explained by the fact that they are more enriched for CpGs and could therefore sequester Uhrf1 more efficiently (Sharif *et al.*, 2016).

Nevertheless, transient derepression of repeats in the germline is beneficial with respect to genome diversification and evolution, in contrast to events in somatic cells which cannot be inherited and could ultimately lead to the development of disease (Crichton *et al.*, 2014). In this regard, the dynamic rate at which DNA methylation can be controlled represents a suitable tool for organisms to employ in order to balance the beneficial and detrimental effects of retrotransposition (Hackett *et al.*, 2012).

Methyl marks outside of these contexts, such as within gene bodies, may also have a function in genome regulation.

Transcription initiated from internal exons, particularly of highly expressed genes, is increased in globally hypomethylated *Dnmt3b*-null mESCs, relative to WT cells. This expression was attributed to the use of alternative promoters in some cases but was principally due to the loss of intragenic methylation that would otherwise prevent spurious PolII entry. Furthermore, these new transcriptional start sites are enriched for Dnmt3b binding and DNA methylation in WT cells, in addition to H3K36me3 upon which Dnmt3b recruitment depends (Neri *et al.*, 2017). This observation is particularly intriguing as the evolutionary ancient organism *Arabidopsis thaliana* exhibits a mosaic methylation pattern, so called as the modification is concentrated across gene bodies and depleted elsewhere (Suzuki and Bird, 2008). Whether in this instance this also contributes to repressing spurious transcription would be indicative of its importance.

Methylation at exons can also influence their splicing into mRNA. This is illustrated during the maturation of T-cell lymphocytes, wherein the binding of CTCF to exon 5 promotes PolII pausing and its subsequent inclusion in CD45 pre-mRNA. DNA methylation is detected across exon 5 in lines irrespective of its inclusion however, it is denser in mature lines within which exon 5 is excluded in the pre-mRNA as CTCF binding is inhibited (Shukla *et al.*, 2011). This raises the possibility that a secondary effect, such as the recruitment of 5mC binding proteins may also influence this pathway.

1.4 Interplay between DNA methylation and the histone code

In vivo, Dnmts must contend with accessing substrates in a nucleosome structure, in which 146bp of DNA are wrapped around an octamer of histones. *In vitro* studies complemented with *in vivo* analyses indicate that these units are refractory to methylation and as such, accessibility is regulated by nucleosome remodelling proteins such as the SNF helicases Lymphocyte-specific helicase (Lsh) and ATRX that likely generate accessible loops of DNA (Geiman and Robertson, 2002; Becker and Workman, 2013; Lyons and Zilberman, 2017). Lsh is particularly important for methylation at LINEs, satellite repeat sequences and ERVS in embryonic and somatic tissues in early development. Interestingly, derepression of satellite and ERVs that is observed in *Lsh*-knockout somatic lines is coincident with DNA hypomethylation, reduction of H3K9me3 and a gain of H3K4me3, suggesting that Lsh recruitment influences multiple levels of epigenetic regulation (Dunican *et al.*, 2013).

As alluded to earlier, there is an interplay between DNA methylation and the so-called histone code. The protrusion of histone tails makes them amenable to a panel of modifications, providing an additional layer of epigenetic regulation. An association has been postulated whereby modified histone tails facilitate or inhibit the methylation of DNA (Du *et al.*, 2015).

Insightful studies using *N. crassa* and *A. thaliana* model organisms have provided evidence of the crosstalk between the epigenetic marks (Freitag and Selker, 2005). In the latter case, a feedback loop between histone and DNA methylation is observed for some DNA methyltransferases (Du *et al.*, 2015). In the mouse, the relationship between the two epigenetic signatures may be mediated on two levels: through a physical interaction between constituent proteins of the pathways and by the mark deposited by one pathway influencing the activity of the other. Furthermore, histone modifications influencing DNA methylation is unidirectional with *N. crassa* yet the literature using mammalian cell lines is potted with inconsistencies. This becomes more apparent in the discussion of Dnmt1 interactors later that have a role in reading histone signatures (Du *et al.*, 2015).

A genome-wide analysis of DNA methylation in mESCs noted that patterns could be explained by vicinal histone methylation marks (Meissner *et al.*, 2008). Generally, promoters of a high CpG density were unmethylated and associated with either house-keeping genes that were expressed and marked by H3K4me3 or silent developmental genes bivalently marked with H3K4me3 and H3K27me3. Meanwhile, low CpG density promoters were mostly methylated, with the exception of those marked by H3K4me2/3. At ICRs, the elements associated with imprinted genes as discussed above, a correlation with H3K9 methylation and DNA hypermethylation was noted. The relationship between these marks and DNA methylation is explored below (Meissner *et al.*, 2008) (see Figure 1.4).

Methylation of H3K9 in mammalian cells is conducted largely by the Suv39 subfamily of 'Suv39, Enhancer of Zeste, Trithorax' domain containing proteins (Dillon *et al.*, 2005; Wu *et al.*, 2010). Despite their differential nuclear localisations, knockout studies and subsequent analyses of the global levels of methylated H3K9 suggests there is some compensation between different Suv39 members, but not sufficient to prevent lethality or adverse effects on viability (Esteve *et al.*, 2006).

The most dominant Suv39 is G9a, which acts in a complex with another transferase GLP, primarily at genic regions. The activity of this heterodimer accounts for the majority of

methylated H3K9, specifically H3K9me and H3K9me2, and is essential for embryonic development (Tachibana *et al.*, 2002). The remainder of H3K9 methylation is accounted for by Suv39h1 and Suv39h2 that act at pericentric heterochromatin and Setd1 which marks ERVs with H3K9me3 (Matsui *et al.*, 2010).

Pericentric heterochromatin is distinguished by minor satellite repeats, which are bound by centromeric proteins. These proteins form complexes that have affinities for Suv39h1/2 and Dnmt3b that will then establish a transcriptionally inert state, mediating deposition of H3K9me3 and DNA methylation respectively. As Dnmt3b recruitment here is not directly dependent on H3K9me3, but rather on the binding of a centromeric protein, it follows that *Suv39h1/2* knockouts do not result in DNA methylation loss at centromeric heterochromatin. However, a correlation is observed at major repeats, at which Dnmt3a can be recruited directly by Suv39h1/2 (Du, Lianna M. Johnson, *et al.*, 2015). Suv39h1/2 appear to propagate H3K9me3 marks deposited in mESCs by Eset which silence intact and potentially functional ERVs and LINEs. The double knockout of *Suv39h1/2* in mESCs can also promote a reduction in DNA methylation at specific LINE-1 families. In neuronal progenitor cells and MEFs, there was shown to be a handover from histone-based mechanisms of repeat repression by Suv39h1/2 and Eset to DNA methylation (Arand *et al.*, 2012; Bulut-Karslioglu *et al.*, 2014).

Although H3K9me2 can independently regulate transcription, it also cooperates with DNA methylation in this capacity.

The accumulation of H3K9me2 occurs rapidly upon exit from naïve pluripotency, following blastocyst implantation and gastrulation, coinciding with an increase in G9a and GLP protein. This can be modelled to an extent *in vitro* through the induction of mESCs to an epiblast-like cell (EpiLC) and eventual epiblast stem cell (EpiLSC) state, respectively (Zylicz *et al.*, 2015). Application of retinoic acid to mESCs *in vitro* promotes differentiation and using this model, pursuant *de novo* methylation in a WT and *G9a*-null mESC line were compared to identify genes whose methylation is driven by G9a. Of the 126 genes identified, *Oct3/4* was selected for further analysis, given its association with maintaining pluripotency. Interestingly, rescue of a *G9a* null line with a catalytically dead protein was sufficient to direct DNA methylation to the *Oct3/4* promoter and prevent its expression upon withdrawal of retinoic acid. This is contrasted with the *G9a* knockout line within which *Oct3/4* was re-expressed in a time dependent manner. However, that the gene was initially silenced demonstrates that histone modifications and DNA methylation are mechanisms that act downstream of other events at

this locus, serving to reinforce, rather than instigate, the silencing of the gene (Epsztejn-Litman *et al.*, 2008).

In a separate study, numerous other genes that were demethylated and subsequently expressed upon the knockout of the *G9a/GLP* heteromer were shown to become remethylated and silenced following the introduction of a catalytically impaired complex. Conversely, H3K9me2 was present at the promoter region of some of these genes in separate *Dnmt1* and *Dnmt3a* and -*b*-null cells lines, within which they were silent. This result would suggest therefore that at these loci, redundant mechanisms operate to silence the genes (Dong *et al.*, 2008; Tachibana *et al.*, 2008). A similar catalytically independent effect of G9a on DNA methylation was also observed at imprinted genes. This study identified the H3K9me1/2 binding ankyrin domain of G9a recruited the *de novo* transferases to deposit DNA methylation (T. Zhang *et al.*, 2016).

In contrast to endogenous sequences, it was shown that the silencing of a virally-transduced leukaemia sequence in mESCs is initially dependent on G9a catalytic activity and Dnmt3a. Directing the suppression after multiple passages however was Eset, provoking the need to consider the influence of transcription upon directing initial epigenetic silencing mechanisms and epigenetic switches when investigating DNA methylation (Leung *et al.*, 2011). Work in other cell line lines has shown that there is an independent recruitment of G9a and DNMTs. For example, *G9a* knockdown or depletion of DNA methylation did not exert a reciprocal effect in HCT116 cells, suggesting early development represents a unique case. The disparity may reflect the epigenetic plasticity of mESCs, compared to the definitive state adopted by committed cells (S. Sharma *et al.*, 2012).

The above observations demonstrate that knockout studies can belie the role of histone methyltransferase proteins in regulating DNA methylation patterns. Overall, it is evident that the G9a protein, in conjunction with GLP, is important for the deposition of DNA methylation through the recruitment of Dnmts directly or indirectly through H3K9me1/2. To delineate the association further, it is apparent that global analyses of *G9a/GLP* knockout and catalytically dead cell-lines should be considered alongside loci specific effects. The validity of these studies would be enhanced if conducted in both pluripotent and differentiated cells as during this transition, transcription, DNA and histone methylation patterns are altered, as does the crosstalk between these.

Methylation of histones can also impede DNA methylation, specifically that of H3K4. Mechanistic studies showed that the addition of methyl groups at this position occludes the interaction between carboxylate groups of the ADD (alpha thalassaemia/mental retardation syndrome X-linked homologue (ATRX)–DNMT3–DNMT3L) domain within Dnmt3l, abolishing an interaction with the histone in the context of H3K4me2/3. Dnmt3l is an important co-factor of Dnmt3a and -b in undifferentiated cells (Ooi *et al.*, 2007). Furthermore, an isolated ADD domain within Dnmt3a was also shown to directly respond to H3K4 methylation, with binding affinities being reduced up to 10-fold *in vitro* (Otani *et al.*, 2009).

DNA methylation can also act instructively by dictating the distribution of H3K27me3. In mammalian tissue, DNA methylation and H3K27me3, modulated by Polycomb Group Repressor complex activity, are generally mutually exclusive at CpG rich sequences. Promoting DNA hypomethylation through DNMT impairment or knockout redistributes H3K27me3. Consequently, gene promoters and enhancer elements are differentially marked compared to hypermethylated controls, with only a minimal increase in global H3K27me3 levels observed. The regions that gain H3K27me3 in a hypomethylated state are associated with high levels of DNA methylation in normal controls and conversely, regions that lose H3K27me3 typically have low levels of DNA methylation in controls. This effect imposes changes on gene expression and those associated with bivalent H3K4me3/H3K27me3 marks in WT cells are typically upregulated, which may be attributed to the loss the H3K27me3 and the subsequent gain of H3K27ac in some cases. In a MEF model at the HoxC cluster, upon loss of H3K27me3 there was accompanying increase in transcription and H3K4me3 across the region that became mostly strongly derepressed. The accumulation of H3K4me3 upon DNA demethylation at the same element however is not observed in mESCs (Reddington *et al.*, 2013; King *et al.*, 2016).

Redistributed H3K27me3 is partly relocalised to gene promoters of a low CpG content and across gene bodies previously devoid of either H3K4me3 or H3K27me3 (Reddington, Pennings and Meehan, 2013). In contrast to the association with H3K9 methyltransferases, the relationship with DNA methylation and H3K27me3 is dependent wholly on the catalytic activity of the Dnmts as rescue of H3K27me3 patterns is only observed with genes that encode catalytically competent Dnmts. Interestingly however, reconstitution of Dnmt3a1 and Dnmt3a2 imparts a greater overall but specific effect, compared to the introduction of Dnmt3b (King *et al.*, 2016). These observations demonstrate the indirect effects of DNA

methylation upon transcriptional regulation and should also be considered when accounting for changes in expression upon global demethylation (Reddington, Pennings and Meehan, 2013).

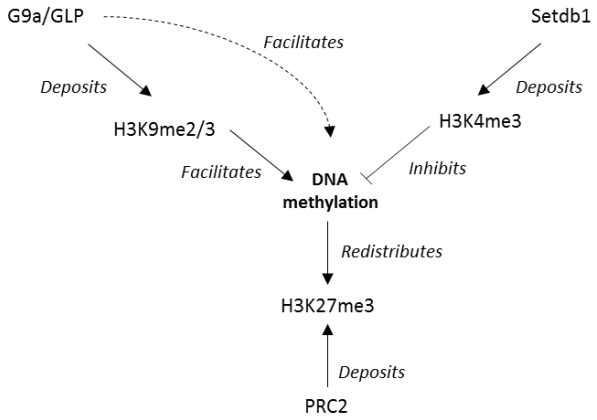


Figure 1.4 – A relationship between the methylation of DNA and histones has been observed. For example, DNA methylation is inhibited at regions vicinal to H3K4me3 whereas there is a strong relationship with H3K9me2/3. In this case, G9a can also recruit the *de novo* methyltransferases independently of its catalytic function. DNA methylation can also act instructively towards H3K27me3 deposition and the marks are generally mutually exclusive on chromatin.

1.5 Dnmt3

The activities of Dnmt3a and -b are the means through which DNA is methylated *de novo* (Lei *et al.*, 1996; Okano, Xie and Li, 1998).

In E7.5 ectodermal tissue, expression of Dnmt3b, which is also uniquely found in mesoderm and endoderm, is considerably greater than Dnmt3a. In E8.5 tissues however, Dnmt3b is detected weakly throughout the embryo and concentrates in the eyes and forebrain, whilst Dnmt3a by comparison is ubiquitous. Double homozygous mutant embryos die before E11.5, which is earlier than the single mutants, thus suggesting some functional redundancy in early development (Okano *et al.*, 1999). Knockout studies in mESCs have shown that whilst Dnmt3a and -b have overlapping targets in some contexts, such as IAPs and several single copy genes, there is specificity at satellite regions for Dnmt3b. Meanwhile Dnmt3a has a preference for SINEs, *Xist* and the *H19* imprinted gene (Okano, Xie and Li, 1998; Okano *et al.*, 1999; Chen *et al.*, 2003; Arand *et al.*, 2012).

Dnmt3b is also particularly enriched across gene bodies. Here, there is an overlap with RNA PolII binding, mRNA and H3K36me3 which are indicative of actively transcribed genes. A possible functional role is to regulate splicing or the silencing of cryptic promoters as discussed earlier (Illingworth *et al.*, 2010; Reddington, Pennings and Meehan, 2013; Baubec *et al.*, 2015).

In mESCs, the major catalytically active isoform of Dnmt3a is Dnmt3a2, which is expressed from an alternative promoter and is ubiquitous in its localisation at chromatin. Upon differentiation, a longer form is expressed, albeit lowly, and is concentrated at heterochromatin in somatic cells. Variants of Dnmt3b arise from alternative splicing and similarly one particular isoform, Dnmt3b1, is more highly expressed in the pluripotent state. This is subsequently undetectable in somatic cells wherein -3b2, -3b4, -3b5 and the catalytically inactive -3b3 isoforms are expressed (Chen *et al.*, 2003; Jurkowska, Jurkowski and Jeltsch, 2011).

Dnmt3a is essential for spermatogenesis throughout which paternal imprints are established. Crosses between WT males and *Dnmt3a* conditionally deleted females, within which the transferase is not expressed in the germline, results in aberrant maternal imprinting, reduced repeat methylation and lethality by E10.5 (Kaneda *et al.*, 2004).

These phenotypes are comparable to the effects of disrupting the Dnmt3l allele, an obligate cofactor of Dnmt3a and -b. This lacks catalytic activity and a portion of the ADD regulatory domain (Bourc'his *et al.*, 2001). In males, Dnmt3l is first expressed in spermatogonial stem cell precursors before birth and assists Dnmt3a in establishing methylation patterns at some imprints and transposable elements. Defects can be inherited in the spermatogonia, triggering aberrant expression of repeats that contributes to meiotic arrest prior to the pachytene stage of spermatogenesis and therefore infertility (Bourc'his *et al.*, 2001; Hata *et al.*, 2002; Kaneda *et al.*, 2004; Webster *et al.*, 2005). Similarly, Dnmt3l is necessary to establish imprints at regulatory regions in oocytes, with offspring exhibiting biallelic expression of unmaintained imprinted genes and lethality by E9.5 (Hata *et al.*, 2002; Bourc'his and Bestor, 2004).

In mESC models cultured in conventional medium, deletion of *Dnmt1* predictably results in a global loss of methylation. Combinatorial knockouts of *de novo* transferases however demonstrate that these are required to top-up patterns that are not faithfully maintained by Dnmt1 alone and therefore offset the dilution of 5mC that would otherwise occur during cell

proliferation. The CpG periodicity of the sequences Dnmt3a and -b methylate is relatively less than those stretches maintained by Dnmt1 alone (Liang *et al.*, 2002; Chen *et al.*, 2003). Another study has shown that high density CpG promoter regions of LINE elements are also dependent on perpetual *de novo* activity (Li, Dai, Suzanne N. Martos, *et al.*, 2015). That recombinant Dnmt3a and -b proteins have equal activity on hemi- and un-methylated substrates *in vitro* strengthens the case that they can act in combination with Dnmt1 (Okano, Xie and Li, 1998). Nascent chromatin capture is a process wherein biotinylated nucleotide analogues are incorporated into replicating DNA and then used as bait to identity interactors in nascent and mature DNA. Data obtained using this technique in HeLa cells supports the observations that latent *de novo* methylation occurs as Dnmt3a was enriched on mature chromatin (Alabert *et al.*, 2014).

Serial passage of *Dnmt3a* and -*b*-null mESCs is therefore associated with diminishing 5mC levels. This also likely impacts the recruitment of MeCP2 as there is a corresponding increase in H4Kac levels. The change in both of these epigenetic marks therefore contributes to the relationship between passage of the *Dnmt3*-null lines and the block in the ability to form teratomas and impaired differentiation capacity (Chen *et al.*, 2003; Jackson *et al.*, 2004). In haematopoietic stem cells, the inverse is observed; Dnmt3a and Dnmt3b are readily detected yet methylation levels are lower than lineage committed hematopoietic cells. When *Dnmt3a* and -*b* are deleted conditionally, cells are capable of differentiation into various lineages, as shown with *in vitro* assays, yet are unable to repopulate the bone-marrow of irradiated mice. Therefore, DNA methylation is unnecessary for lineage commitment from differentiated multipotent HSCs but is indispensable for self-renewal. Of note is that singularly deleting a *de novo* transferase had no effect, providing evidence of redundancy in this system (Tadokoro *et al.*, 2007).

That DNA methylation reaches a nadir in *Dnmt3a*, *Dnmt3b* -/- mESCs suggests that a minimum amount of DNA methylation is required before Dnmt1 recruitment falters genome-wide. As differentiated cells have lower Dnmt3a and -b levels compared with mESCs cultured in traditional fetal calf serum (FCS)-based media, this would suggest the targeting of Dnmt3a and -b is highly efficient to minimise a progressive loss in DNA methylation in somatic cells (Liang *et al.*, 2002; Baubec *et al.*, 2015).

De novo methylation during spermatogenesis in rodents is also mediated by Dnmt3c, a Dnmt3b paralogue that contains all the functional methyltransferase motifs and the N-

terminal domain required for H3K4 binding. This specifically methylates the promoters and supresses activity of young and transcriptionally active LINEs and ERVs, which partially overlaps with the elements derepressed in *Dnmt3l*-null germ cells (Barau *et al.*, 2016).

In addition to the influence of histone modifications, *de novo* methylation anticorrelates with CpG density and transcription factor occupancy of a defined region within the promoter of a gene, even in the absence of active transcription (Lienert *et al.*, 2011). Data also supports the notion that *de novo* methylation may occur in a feed-back loop whereby pre-existing methylation at an ERV increases the likelihood that it will be maintained in a clonal population whereas contrarily, intermediate levels result in a stochastic gain or loss of 5mC. Whilst this was observed in Murine Erythroleukemia (MEL) cell lines, mESCs in the same study were shown to methylate the ERV regardless of the starting levels of 5mC (Lorincz *et al.*, 2002).

1.6 Dnmt2

Before discussing Dnmt1 in detail, it is worth bringing to attention Dnmt2. This is present in almost all eukaryotes, including humans and mice. Structurally it lacks a large N-terminal regulatory region but contains all the characteristic sequences of a methyltransferase catalytic domain (Jurkowski and Jeltsch, 2011). However, the substrates for Dnmt2 are tRNAs -Asp, - Gly and –Val at the C38 position and methylation stabilises the structures. In *Drosophila*, Dnmt2 is the only Dnmt-like protein and a role in DNA methylation has not been completely excluded yet (Lyko, Ramsahoye and Jaenisch, 2000; Kunert *et al.*, 2003; Reviewed: Jeltsch *et al.*, 2017). In isolation, the knockout of *Dnmt2* in mice is not sufficient to generate a strong developmental phenotype with a reduced rate in the synthesis of proteins containing poly-Asp peptide runs being one of the few aberrations found so far (Jurkowski and Jeltsch, 2011; Jeltsch *et al.*, 2017). However, if the tRNA-Leu methyltransferase *NSun2* is also targeted, pups are born smaller than WT, with reduced rates of protein synthesis, an underdeveloped cerebral cortex and at ratios significantly less than expected by Mendelian inheritance (Tuorto *et al.*, 2012; Jeltsch *et al.*, 2017).

An important role of Dnmt2 in modifying heritable small non-coding RNAs (sncRNAs) derived from tRNAs in sperm has also recently come to light. If a donor male is maintained on a high fat diet, glucose intolerance is inherited in the offspring. Notably, sperm from *Dnmt2*-null

mice does not produce this result, in which there is likely a disruption in the pool of sncRNAs, due to the impact on tRNA stability and a lack of sncRNA methylation (Zhang *et al.*, 2018).

1.7 Dnmt1

Prior to mitosis, nascent DNA must be methylated in accordance with the pattern on the parental DNA strand and maintained as such during subsequent cell divisions. This is catalysed by Dnmt1 in conjunction with cofactors of varying necessity that ensure the inheritance of the methylation pattern. Although a compartmentalised model was presented initially, the *de novo* methyltransferases as discussed above also contribute during this process. This is necessary as the fidelity of Dnmt1 in mESCs is estimated to be approximately 98% (Jackson *et al.*, 2004).

Displaying 86% sequence identity, it is unsurprising that crystal structures for mDnmt1 and hDNMT1 in conjunction with a DNA substrate suggests the same mode of action. Overlays of the bacterial M.Hha-1-DNA complex are also comparable for the majority of the ten conserved sequence motifs in the catalytic domain that define a methyltransferase (Takeshita *et al.*, 2011; Z.-M. Zhang *et al.*, 2015).

In vivo, Dnmt1 has to recognise the appropriate substrate while interacting with the methylgroup donor AdoMet, before forming and then subsequently breaking a covalent bond with the cytidine ring of a nucleotide upon transfer of the methyl group (Takeshita *et al.*, 2011). Once engaged with chromatin, the enzyme operates in a processive manner, transferring methyl groups to multiple hemimethylated sites without dissociating from the substrate after each individual reaction (Fatemi *et al.*, 2001). Fidelity is increased at regions dense in CpG dinucleotides and is compensated for at sparser regions by the co-factors discussed below (Goyal, Reinhardt and Jeltsch, 2006). Replication foci arising at the earliest stage of the S phase of the cycle show an enrichment of Dnmt1, yet it becomes more concentrated at decondensed heterochromatin during the latter stages (Schermelleh *et al.*, 2007).

In mESCs and other cell types, whilst the majority of methylation patterns are restored within 1 hour post-replication, it takes a further 3 hours for all the appropriate marks to be established to the same extent as the parental cell (Charlton *et al.*, 2018). This suggests a replication-uncoupled mechanism of Dnmt1 activity may also be operative, in addition to the

contributions of the *de novo* methyltransferases. Nevertheless, the modification status of neighbouring CpG dinucleotides correlate strongly, supporting the notion that Dnmt1 acts in a processive manner. This can give rise to a heterogenous population of cells that are present within different phases of the cell cycle and with various extents of methylation. Active transcription factor binding sites are preserved in a partially methylated state thus, demonstrating the influence transcription imposes on the methylome (Liang *et al.*, 2002; Charlton *et al.*, 2018).

Dnmt1 is a large multi-domain enzyme of 1620 amino acids in the mouse and 1616 in humans (Jurkowska, Jurkowski and Jeltsch, 2011). The structure will now be discussed in a manner starting at the C-terminal catalytic domain through to the N-terminal regulatory domain (see Figure 1.5).

Figure 1.5 – Schematic of the functional domains of Dnmt1. The role of each domain is discussed in the main text in detail.

1.7.1 Catalytic domain

The methyltransferase domain of Dnmt1 is composed of a catalytic core and a target recognition domain (TRD) that confers an intrinsic affinity for hemimethylated sites, calculated to be over 40-fold relative to unmethylated nucleotides (Hermann, Goyal and Jeltsch, 2004). Engaging with double stranded DNA promotes a conformational change of this domain (Ye *et al.*, 2018).

Mutation of a series of evolutionarily conserved residues that comprise a hydrophobic pocket in the TRD that envelopes the methyl group on the parental strand, and others that stack with the base, can abolish methyltransferase activity or the specific affinity for hemimethylated substrates (Takeshita *et al.*, 2011; Kanada *et al.*, 2017). On the opposing unmethylated strand, a catalytic cysteine points towards the target cytosine residue that is flipped out of the DNA strand upon binding of AdoMet, and this is further stabilised by interactions with highly conserved residues that contact the base and the phosphate backbone (Song *et al.*, 2012). Crystal structures of human DNMT1 show the catalytic residue

is poised for catalysis prior to an interaction with the methyl-donor AdoMet, whereas a conformational change is induced in mDnmt1 upon interacting with this co-factor (see Figure 1.7) (Z.-M. Zhang *et al.*, 2015).

1.7.2 GK, BAH & CXXC Domains

A stretch of glycine-lysine (GK) repeats is situated between the catalytic and N-terminal regulatory domains, a feature that is conserved amongst all eukaryotic homologues (Yarychkivska, Tavana, *et al.*, 2018). This motif is highly flexible, rendering its resolution in crystal structures difficult. Nevertheless, substitution of the lysine residues with arginine, which prevents their acetylation, imparts *de novo* methyltransferase activity at paternally imprinted regions when the construct is used in *Dnmt1* -/- mESC rescue experiments. This raises the possibility that selective deacetylation of Dnmt1 in spermatogonia imbues it with an ability to establish imprints (Yarychkivska, Shahabuddin, *et al.*, 2018). This also prompts the question as to whether specific insects and fungi that seemingly only have a maintenance enzyme also have integral mechanisms that promote a switch to a *de novo* methylation capacity (Bewick *et al.*, 2016; Catania *et al.*, 2017).

The highly conserved tandem bromo adjacent homology (BAH) domains 1 and -2 are also demonstrably important for Dnmt1 activity. These are linked by an alpha-helix and when folded, adopt a dumb-bell configuration with BAH1 orientated closely to the catalytic domain. Found in a wide range of proteins with a chromatin remodelling function across different species, deletion of the BAH domains significantly impairs Dnmt1 recruitment and its capacity to maintain DNA methylation genome wide in mESCs, possibly owing to the absence of an essential intermolecular interaction (Yang and Xu, 2013; Yarychkivska, Shahabuddin, *et al.*, 2018).

A preparation of the isolated CXXC Zn finger DNA-binding domain has been described as possessing an affinity for both methylated (Fatemi *et al.*, 2001) and unmethylated (Pradhan *et al.*, 2008) DNA, indicating that Dnmt1 has multiple potential DNA binding sites. Nevertheless, others have shown that this is not essential for the loading of Dnmt1 onto chromatin (Schneider *et al.*, 2013). This domain was initially posited to act in an autoinhibitory manner from resolved crystal structures of truncated proteins, contacting unmethylated cytosine residues to block methylation by the catalytic domain. However, this was not confirmed in other studies using full-length Dnmt1 in crystal structures or *in vitro* assays (Song *et al.*, 2011; Takeshita *et al.*, 2011; Bashtrykov *et al.*, 2012).

The N-terminal domains that are by far the subject of most intensive focus are the proliferating cell nuclear-antigen (Pcna) binding domain (PBD) and replication foci-targeting sequence (RFTS) because of their functions in recruiting co-factors necessary for the loading of Dnmt1 onto chromatin.

1.7.3 Pcna Binding Domain and RFTS

Despite the affinity of the Dnmt1 catalytic domain toward hemi-methylated substrates, this property alone is insufficient to maintain methylation patterns *in vivo.* There is a dependency on additional recruitment factors that likely act cooperatively to target Dnmt1 to chromatin at the appropriate stage of the cell-cycle. In this regard, Dnmt1 is assisted by Pcna and Uhrf1 that bind to specific sequences in its N-terminal regulatory domain. The importance of this terminus as a whole is also underscored by the detection of domain-domain interactions with the C-terminus and the inability of the isolated catalytic domain to methylate substrates *in vitro*. This suggests that a role of the N-terminal domain is to maintain Dnmt1 in an appropriate conformation (Fatemi *et al.*, 2001). During proteolysis, fragments of the C-terminal domain will therefore be unable to aberrantly methylate DNA if they associate with chromatin (Jurkowska, Jurkowski and Jeltsch, 2011).

The RFTS to which Uhrf1 binds also serves an autoinhibitory function, with mutants impaired for an interaction between this domain and the catalytic region displaying greater activity *in vitro* (Bashtrykov, Rajavelu, *et al.*, 2014). When unbound to chromatin, the RFTS is inserted deeply into the catalytic pocket. This conformation is sustained by direct electrostatic interactions and hydrogen bonding, along with a linker helix between the CXXC and BAH1 domain contacting both the RFTS and catalytic domain. The trigger for the unfolding of this structure is likely promoted by direct interaction with the SET and RING associated (SRA) DNA binding domain of Uhrf1 (Bashtrykov, Jankevicius, *et al.*, 2014; Bashtrykov, Rajavelu, *et al.*, 2014). A range of neurological conditions associated with Dnmt1 mutations have been mapped back to this region and it is conceivable that the dysregulated enzymatic activity is due to impaired recruitment (Z.-M. Zhang *et al.*, 2015).

There is also a report that this domain mediates a homotypic dimerization and as it is particularly hydrophobic in nature, this property is likely the driving intermolecular force. Overall the necessity of this remains relatively unexplored and the temporal association of two Dnmt1 proteins requires investigation (Fellinger *et al.*, 2009).

It was proposed that the loading of Dnmt1 hinges on Pcna in early to mid S phase, switching to Uhrf1 at latter stages wherein densely methylated and repetitive heterochromatin is replicated and therefore multiple hemi-methylated sites are presented for renewal. Partial deletion of the RFTS domain (459-501), which Uhrf1 binds, results in a weaker association to foci (Takebayashi *et al.*, 2007; Schneider *et al.*, 2013) and deletion of Uhrf1 results in global hypomethylation, impacting patterns in both euchromatin and heterochromatin (Bostick *et al.*, 2007; Sharif *et al.*, 2007). The differentiation potential of *Uhrf1*-null mESCs is also equivalent to *Dnmt1*-null mESCs (Bostick *et al.*, 2007). This suggests Uhrf1 is the main regulator of Dnmt1 activity.

Alternative splicing can give rise to a functional variant of Dnmt1 that lacks the PBD. This displays a ~2-fold lower efficiency in recruitment to replication forks and work in HCT116 cells suggests that this equates to a 20% reduction in global methylation. The cells remain viable and healthy *in vitro* until further knockdown of the truncated transcript, indicating PCNA is not essential for the maintenance of DNA methylation (Schermelleh *et al.*, 2007; Spada *et al.*, 2007). This result can also be recapitulated in mESCs where Dnmt1 defective in Pcna binding can substantially restore methylation in *Dnmt1* -/- lines. This suggests that the interaction is not strictly essential across multiple cell types and its involvement is necessary but limited to the recruitment during the early to mid-stages of the S phase in the cell cycle only (Schermelleh *et al.*, 2007; Spada *et al.*, 2007; Takebayashi *et al.*, 2007). These observations also uncouple DNA methylation from replication, a conclusion enforced experimentally by immobilising Dnmt1 on chromatin which can still permit DNA replication to proceed (Spada *et al.*, 2007).

1.8 Uhrf1

The multi-domain composure of Uhrf1 would feasibly enable it to assist Dnmt1 in processing hemimethylated sites, linking DNA methylation with histones through multivalent interactions. Uhrf1 contains the H3K9me2/3 binding tandem tudor domain (TTD) and plant homeobox domains (PHD), in addition to a hemimethylated CpG binding SRA domain (see Figure 1.6). Homologues found in *A. thaliana* and insects demonstrably perform similar functions but they are not necessarily dependent on the reading of histone marks (Bostick *et al.*, 2007; Du *et al.*, 2015; Bewick *et al.*, 2016). Furthermore, Uhrf1 can also bind directly to

unmodified H3, albeit inefficiently, and this interaction is enhanced by hemimethylated DNA (Fang *et al.*, 2016).

mESC lines heterozygous for WT and catalytically dead *Dnmt1* alleles display a normal cell-cycle dependent distribution of Dnmt1, suggesting that the residual methylation maintained by the WT protein is sufficient for Uhrf1 to bind the hemimethylated DNA and recruit both Dnmt1 variants. This is supported by the diffuse localisation of Dnmt1 in a heterozygous null/catalytically dead embryonic somatic cells which are globally hypomethylated. In mESCs with this genotype however, Dnmt1 localised to pericentric heterochromatin throughout the S phase (Takebayashi *et al.*, 2007). As *de novo* methyltransferases are expressed in these cells, the amount of DNA methylation that is maintained may have been sufficient for Dnmt1 recruitment by Uhrf1, coupled to an interaction with histone marks.

Due to the mitotically arrested state of oocytes, evidence of Dnmt1 activity is minimal. It was therefore surprising that Uhrf1 knockdown results in a ~20% loss DNA in methylation, specifically across moderately methylated regions. This suggests that Uhrf1 in certain contexts can also influence *de novo* methylation (Maenohara *et al.*, 2017). As *de novo* methyltransferases have an intrinsic preference for CpGs flanked by a purine at the 5' end and a pyrimidine at the 3' end , Uhrf1 may recruit Dnmt1 to methylate the opposing CpG that does not have favourable neighbouring bases (Ravichandran, Jurkowska and Jurkowski, 2018).

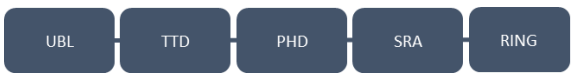


Figure 1.6 – Schematic of the functional domains of Uhrf1. The role of each domain is discussed in the main text in detail.

1.8.1 SRA domain

The SRA domain of Uhrf1 is a concave structure which undergoes rearrangement upon binding of 5mC to enclose the opposing DNA strands. The methylated cytosine within the symmetrical dyad is flipped out and into a charged patch within the concave, stacking with highly conserved residues. Other residues of the SRA domain enter the vacated space and interact with neighbouring bases, including the guanine base previously bound to the flipped-

out methylated cytosine. Loops that emerge from the conformational change may stabilise the interaction as they contact both minor and major grooves of the DNA. Despite all of this, the DNA retains its overall structure when engaged with Uhrf1. The configuration of the SRA regions and subsequent extensive network of interactions between the base and the pocket have been presented as the means through which hemimethylated patterns are discriminated from fully methylated dyads and other nucleotides, thus conferring specificity (Arita *et al.*, 2008; Avvakumov *et al.*, 2008; Hashimoto *et al.*, 2008). Dnmt1 successively flips out the corresponding cytosine on the nascent strand which likely occurs, given the potential for steric clashes otherwise, upon Uhrf1 disembarking from the methylated cytosine (see Figure 1.7) (Hashimoto *et al.*, 2008). Intriguingly, given the inhibitory effect on activity, 5-hydroxymethylcytosine (5hmC) is also bound by the SRA domain of Uhrf1 with an affinity comparable to that of 5mC interactions (Frauer *et al.*, 2011). The consequences of this on the maintenance of DNA methylation however remains to be investigated.

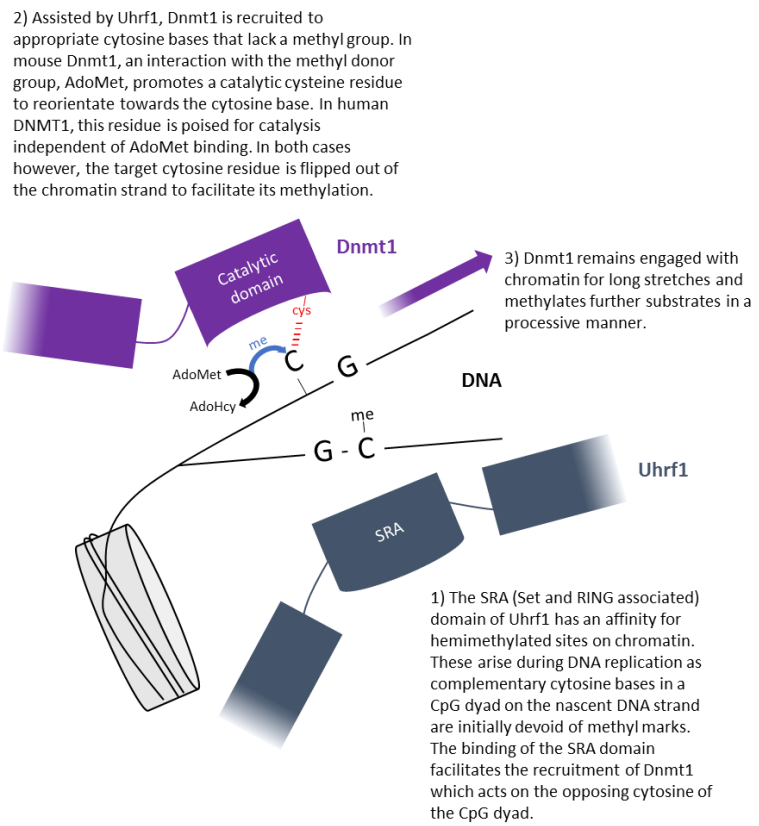


Figure 1.7 – An overview of Dnmt1 and Uhrf1 facilitated maintenance DNA methylation. See main body of text of references.

At the molecular level, binding of hemimethylated DNA induces a conformational change in Uhrf1 whereby blockage of the TTD by a spacer region situated between the SRA and RING domains is alleviated. This enables a more favourable interaction between PHD-TTD and histones and for the spacer region to facilitate SRA mediated DNA binding otherwise as shown *in vitro*, histone modifications do not enhance an interaction between hemimethylated DNA and Uhrf1 (Fang *et al.*, 2016).

1.8.2 TTD and PHD domains

Introduction of a point mutation (Y188A, human) in the aromatic cage structure of the TTD renders Uhrf1 incapable of binding H3K9me2/3. This mutant was unable to rescue DNA methylation in *Uhrf1*-null HeLa cells (Rothbart *et al.*, 2012). Furthermore, mESCs deficient in DNA methyltransferases display a diffuse Uhrf1 localisation pattern, which would initially suggest that both marks, H3K9me2/3 and 5mC, are required for proper recruitment (Sharif *et al.*, 2007; Rothbart *et al.*, 2012). Direct evidence for the requirement of existing DNA methylation to recruit Dnmt1 comes from a study in which WT *Dnmt1* was transfected into TKO cells and yet it remained diffuse and this was partially rescued with the reintroduction of a *de novo* transferase (Sharif *et al.*, 2007; Takebayashi *et al.*, 2007).

The TTD also robustly interacts with DNA ligase 1 (Lig1) and this is abolished upon mutation of residues that accommodate H3K9me2/3 (human Y188A/Y191A). Further inspection of the LIG1 domain that mediates the interaction showed it bears a TARK motif that is found in H3 tails and methylation of the lysine residue within this by G9a/GLP mediates the interaction with Uhrf1. Methylated Lig1 dependent co-localisation with Uhrf1 was confirmed in mESCs which had concordant effects on Dnmt1 too, demonstrating they form a complex (Ferry *et al.*, 2017).

The study from which this model was developed was fuelled by the disparity between convenient observations that Uhrf1 has a 10-fold more affinity to hemi-methylated DNA and can bind H3K9me2/3, against the sheer potential number of hemimethylated CpG sites during cell division: 25 million methyl cytosines per human cell and 30 million nucleosomes, with 50% bearing H3K9me2/3. Whilst this Lig1 model finesses how Dnmt1 is recruited, it is complicated by the fact that *Lig1* -/- mESCs lines have no detectable loss of methylation, in contrast to mutants lacking the residues methylated by G9a/GLP. Furthermore, analyses of methylation levels of the mutated *Lig1* cells which give striking effects on Uhrf1 recruitment,

still have ~30% more methylation than *Uhrf1*-null cells. This suggests that an alternative DNA ligase, such as Lig3, may substitute (Ferry *et al.*, 2017).

What requires reconciling for all models implicating histone or histone mimic modifications in Uhrf1 recruitment is that mice homozygous for a knock-in point mutation that renders Uhrf1 deficient for H3K9me2/3 binding are viable and fertile (Zhao *et al.*, 2016). Conversion of tyrosine 187 and proline 188 to alanine in mUhrf1, equivalent to those used in the study cited above (Rothbart *et al.*, 2012) was sufficient to abolish H3K9me2/3 binding, however this only equated to an average of 10% less methylation across the genome in multiple tissues tested. This had no subsequent effect on H3K9me3, H3K27me3 or histone acetylation. Given the abundance of H3K9me2/3 over repetitive elements, it was surprising that DNA methylation levels were comparable to the global average loss. The reduction was accounted for by an inability of the mutated Uhrf1 to bind nucleosomes and utilise H3K9me2/3 to facilitate recruitment of Dnmt1 at demethylated sites at the dyads of nucleosomes (Zhao *et al.*, 2016). Furthermore, a subset of genes, attention to which was drawn earlier, whose expression is regulated expressly by DNA methylation are also depleted of H3K9me2/3 (Hackett *et al.*, 2012). Collectively, this suggests DNA methylation can be maintained largely independently of H3K9me2/3 in differentiated tissue.

On the other hand, ablating H3K9me3 levels through knockdown of *Suv39h1/2* had no significant impact on Uhrf1 recruitment in NIH-3T3 cells. The effect on DNA methylation was not investigated in this case and recruitment may still have been achieved through binding to H3K9me2 marks deposited by alternative histone modifying enzymes, such as G9a/GLP (Papait *et al.*, 2008).

The implications of these studies are that *in vivo*, the predominant substrate for Uhrf1 binding is hemimethylated DNA, mediated through the SRA domain. Therefore, why Dnmt1 is responsive to histone modifications appears difficult to reconcile, particularly under the burden of evidence that healthy mice are born if the Uhrf1-H3K9me2/3 interaction is disrupted (Zhao *et al.*, 2016). Could it be therefore that mESCs are a poor reflection of what truly occurs developmentally regarding epigenetics? The transient moment for which the state exists *in vivo* and the long-term culture of mESCs could be the source of the discrepancy. Global DNA methylation analyses and biochemistry would enhance our understanding of how strict the dependency on histone modifications or the proteins that conduct these are

in setting up or maintaining 5mC patterns in mESCs and somatic cells (Dong *et al.*, 2008; Zhao *et al.*, 2016).

1.8.3 RING and UBL domains

As the name suggests, Uhrf1 also has ubiquitination activity in the form of a RING domain and is therefore both a reader and writer of histone marks. The substrate in mammalian lines was identified as H3K18, a prerequisite for the localisation of Dnmt1 to chromatin (Qin *et al.*, 2015). Further work has revealed additional Uhrf1-dependent mono-ubiquitinated sites to which Dnmt1 binds in coordination with similar affinities: H3K14 and H3K18, K14 and K23 and at K18 and K23 (Ishiyama *et al.*, 2017).

Although the RING domain is not required for Uhrf1 localisation and interaction with Dnmt1 per se, when the ubiquitination function is defective, the recruitment of Dnmt1 is impaired and consequently methylation patterns are only marginally maintained during S phase (Nishiyama *et al.*, 2013; Qin *et al.*, 2015). The identification of a unique and conserved ubiquitin interacting motif (UIM) (mouse 380-399) in Dnmt1 provides evidence of a direct relationship. Interestingly, when this is deleted or mutated, there is an increased affinity for core histones. As this domain maps to the RFTS that also serves an autoinhibitory function, binding to H3K18 may be the trigger for the displacement from the catalytic domain. This is supported by *in vitro* studies demonstrating augmented Dnmt1 activity in the presence of H3Ub2 (Qin *et al.*, 2015; Ishiyama *et al.*, 2017). Furthermore, the ubiquitin like (UBL) domain of Uhrf1, which in isolation can stimulate Dnmt1 *in vitro* by 2-fold, structurally resembles ubiquitin. Deletion of this segment also negatively impacts Dnmt1 recruitment in mESC models (T. Li *et al.*, 2018) and is critical for the E3 ligase activity. This is due to its role in binding to the E2 ligase UbcH5a and facilitating the transfer of a ubiquitin molecule UbcH5a bears to the RING domain of Uhrf1 (Foster *et al.*, 2018).

Ubiquitination is also dependent on binding of the PHD domain. Reportedly, this is due to recognition of H3R2 however, given its proximity to the ubiquitinated residues and a resolved crystal structure showing the residue within a lobe formed by RFTS, further work is required to understand the role (Qin *et al.*, 2015; Ishiyama *et al.*, 2017).

The inclusion of the ubiquitination component to the Dnmt1 recruitment model presents an intermediate state that accounts for the fact that a stable interaction between Uhrf1 and Dnmt1 is not necessarily found unless the components are overexpressed (Nishiyama *et al.*,

2013; Lu *et al.*, 2015) (see Table 1.1). The identification of the UIM also raises the possibility that Dnmt1 can interact with the polycomb system. Indeed, Dnmt1 was found to bind to K119Ubq that is deposited by RING1A/B, components of PRC1. Dnmt1 can also interact directly with Ezh2 of PRC2, which may facilitate its recruitment (Qin *et al.*, 2015).

Taken together, one could posit a model in which Uhrf1 is localised to chromatin, primarily dictated by hemimethylation, which then sets up a more permissive state for Dnmt1 activity and facilitates its recruitment through H3K119Ubq. Experimental evidence supports a minor role of topoisomerase II-α (TopII-α) in regulating Dnmt1 activity at pericentric heterochromatin, albeit indirectly through an interaction with Uhrf1. The resolution by TopII-α of precatenanes produced after DNA replication may present a more favourable substrate for Dnmt1 (Lu *et al.*, 2015) (see Figure 1.8).

Table 1.1 – Approaches used to detect Dnmt1–Uhrf1 interactions

Bait	**Cell line**	**Reference**
Biotin-labelled Uhrf1	mESCs	(Sharif *et al.*, 2007)
CBD-DNMT1	HEK-293T	(Bostick *et al.*, 2007)
GFP-UHRF1 /His-DNMT1	*Dnmt1*-/- mESCs	(Qin *et al.*, 2015)
HA-FLAG-UHRF1	HeLa	(Ferry *et al.*, 2017)
Endogenous DNMT1	HeLa	(Felle *et al.*, 2011)

1.9 Dnmt1 interactors

The N-terminus of Dnmt1 is a platform for the binding of numerous transcriptional repressors. This has led to enquiries into the extent to which the phenotype of *Dnmt1*-null mice is dependent on DNA methylation per se or functions independent of this. Although the same outcome, lethality at E10.5, manifests in mouse models with a catalytically dead protein, this severe phenotype would still obscure other functions (Takebayashi *et al.*, 2007).

The N-terminus of Dnmt1 can bind HDAC1, HDAC2 and Dnmt1-associated protein 1 (DMAP1). These contribute to transcriptional repression either directly, or in the case of the latter, through interaction with other factors such as TSG101 (tumour susceptibility gene 101)

(Rountree, Bachman and Baylin, 2000; Fuks *et al.*, 2000). DMAP1 is critical for pre-implantation development in its role as a TIP60/p400 complex component, maintaining ESC identity and differentiation potential (Fazzio, Huff and Panning, 2008), however this function appears to be mediated independently of Dnmt1. Conversely, levels of DMAP1 are important in the maintenance of imprints in early development (Rountree, Bachman and Baylin, 2000).

The endogenous interaction of retinoblastoma (Rb) protein and E2F1 with the N-terminal domain of DNMT1 suggests that this portion can serve as a binding platform for multiple synergistically acting transcriptional repressors. This was shown to indeed be the case when HDAC inhibitors or a catalytically dead DNMT1 were only able to partially relieve inhibition of a reporter transgene. Likewise, when DNMT1 was co-transfected with Rb, there was repression of an E2F1 responsive ARF reporter. The use of an artificial E2F1 responsive reporter however demonstrated that DNMT1 alone was insufficient to modulate transcription and that Rb was required to initiate repression, which was potentiated by increasing DNMT1 amounts. Taken together, this suggests that DNMT1 could be specifically targeted to genes with E2F1 binding sites and that whilst Rb initiates repression in this context, silencing is ensured by DNMT1 activity or a dependent recruitment of additional repressors such as HDACs 1 and -2 (Robertson *et al.*, 2000). Similar interactions with transcription factors have also been reviewed and represent a highly specific mode of gene regulation (Hervouet *et al.*, 2018).

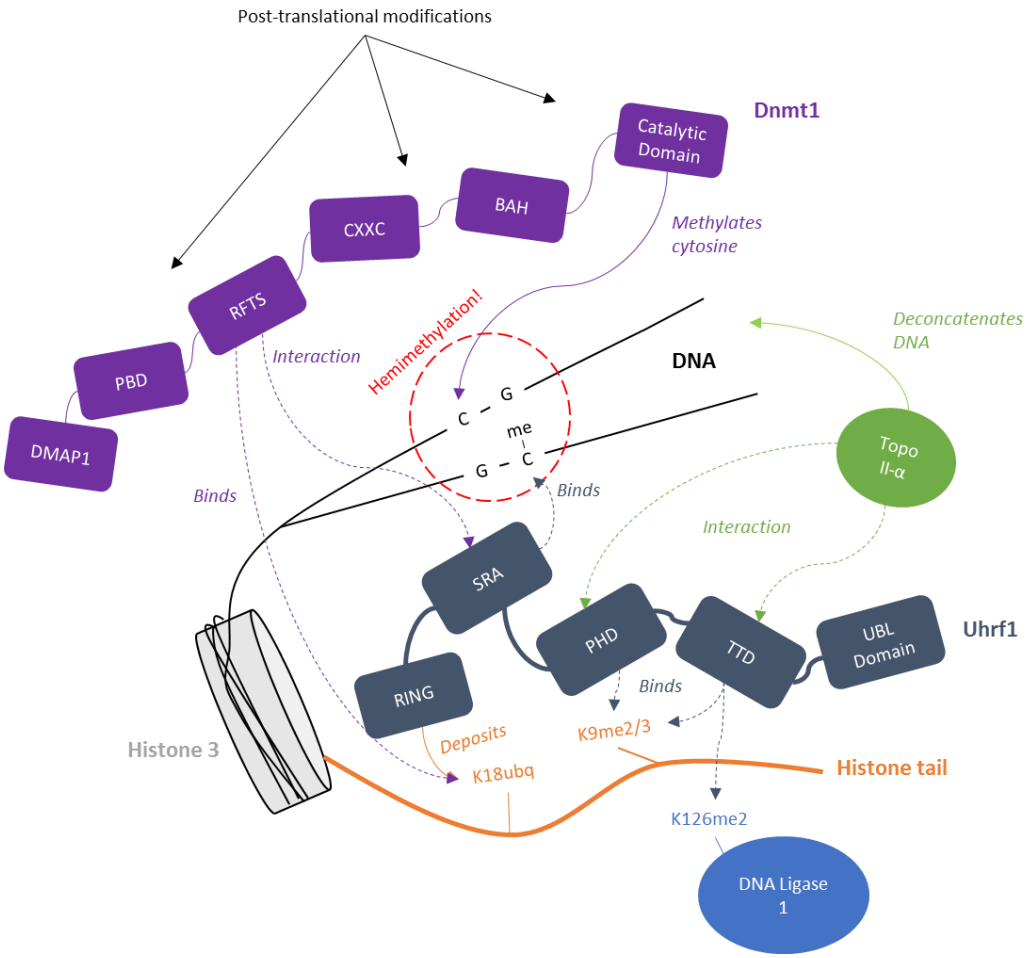


Figure 1.8 - The recruitment of Dnmt1 to hemi-methylated sites and its activity in restoring a fully methylated CpG dyad is dependent on Uhrf1 for multiple reasons. The description that follows is not necessarily ordered in the sequence of events that occur *in vivo* as these have yet to be fully established. Dnmt1 adopts a conformation wherein the RFTS is inserted deep into its own catalytic domain. Upon interacting with the methylation binding SRA domain of Uhrf1, the RFTS is prompted to be removed. This presumably facilitates the interaction of the RFTS with ubiquitinated H3K18Ubq marks that are deposited by the RING domain of Uhrf1. Uhrf1 itself is localised to chromatin through a network of interactions between K9me2/3, mediated by PHD and TTD domains, and by the binding of the SRA domain to methylated sites. Other ancillary proteins assist in the loading of Dnmt1 onto chromatin. For example, Topoisomerase II-α (Topo II-α) interacts with Uhrf1 and likely deconcatenates DNA to present it in a more favourable conformation for DNA methylation. Methylated DNA ligase 1 meanwhile can direct Uhrf1 to chromatin at H3K9me2/3 depleted regions. Furthermore, this whole process occurs under the influence of post-translational modifications which can affect protein stability and activity. See main body of text for references and abbreviations. Adapted from Qin *et al.*, 2015.

Staining for heterochromatin protein 1- β (HP-1 β), and its binding site H3K9me2/3 in *Dnmt1* homozygous null and *Dnmt1*-null/catalytically dead mESCs would suggest that the localisation of HP-1 β is dependent on Dnmt1 recruitment to chromatin. Left unanswered however was whether it is dependent on Dnmt1 per se or on DNA methylation as catalytically dead Dnmt1 is unable to bind efficiently to chromatin (Takebayashi *et al.*, 2007). There is also evidence that the chromodomains of all three mammalian isoforms of HP-1 (α, β, γ) co-purify with DNMT1 N and C-term domains in HCT116 lysates. *In vitro* assays suggested that DNMT1 can increase HP-1 binding to an immobilised DNA substrate independent of its catalytic activity. This effect is reciprocated as the incubation of DNMT1 with HP-1 proteins promoted a 3-fold increase in the rate of substrate methylation. Furthermore, in more complex nucleosome arrays, the necessity of G9a dependent histone methylation was shown in order for HP-1 to be recruited to histones and to subsequently promote DNMT1 activity (Smallwood *et al.*, 2007).

Direct relationships between Dnmt1 and G9a have also been explored in cell lines and with *in vitro* assays. Knockdown of *Dnmt1* in HeLa cells impaired the loading of G9a onto chromatin and reduced total H3K9me2 levels by 50%, compared to a 65% reduction upon *G9a* knockdown. An *in vitro* DNMT1 activity assay within which full length G9a and DNMT1 were incubated, indicated that the association increases the rate of DNA methylation. Furthermore, in the absence of G9a in a colon cancer cell line, the methylation of a selected sequence was reduced by 60%. The reciprocal experiment, to determine whether Dnmt1 promotes G9a activity, revealed only a relatively modest positive effect *in vitro* (Esteve *et al.*, 2006).

This raises the possibility that G9a may have catalytically independent roles in maintaining DNA methylation, as has been shown for Dnmt3a and -b (Dong *et al.*, 2008). The reciprocity model however does not translate wholesale to mESCs as *Dnmt1*-null and TKO cells have normal levels of H3K9me2/3 (Tsumura *et al.*, 2006; Takebayashi *et al.*, 2007). This relationship also contradicts that discussed earlier demonstrating G9a is dispensable for DNA methylation in other somatic cells (S. Sharma *et al.*, 2012).

Immunoprecipitations have also detected a DNMT1-SUV39H1 interaction, although this occurred post replication. The chromodomain of SUV39H1 can bind H3K9me2, which may then be modified further to H3K9me3 (Esteve *et al.*, 2006). It would be of interest to

determine to what extent this SUV39H1 recruitment can account for post-replicative DNA methylation maintenance.

1.9.1 Transcriptional control of Dnmt1

Multiple signalling effectors have been implicated in the transcriptional regulation of DNMT1. This includes Ras which can be accounted for by the presence of c-Jun dependent enhancers and AP-1 binding sites amongst the four identified transcription initiation sites of *DNMT1*. An interplay with the inhibitory effects of p53 and Sp-1 binding has also been described, in addition to inhibitory roles of microRNAs acting at the 3'UTR and protein coding regions (Kar *et al.*, 2012). The 3'UTR of *DNMT1* mRNA is also a site at which AUF1/hnRNP D protein binds, as demonstrated in HeLa, HEK-293 and T24 cells. Overexpression of AUTF1 reduced *DNMT1* mRNA stability however, as this was accompanied by impaired cell-growth, this effect could have been non-specific. Conversely, AUF1 knockdown led to increased DNMT1 protein. As levels of AUF1 anticorrelate with *DNMT1* mRNA, this is an attractive candidate that accounts for the cell cycle dependent regulation of DNMT1 levels in particular tissues (Torrisani *et al.*, 2007).

Furthermore, the human DNMT1 protein was shown to be susceptible to inhibition by RNAs, namely the long non-coding RNA (lncRNA) asCEBPA-2HPE and micro-RNA (miRNA) miR-155-5p, amongst others. These likely act as competitive inhibitors against the hemi-methylated substrate and are most potent when present in a G-quadruplex form. Whilst the effect was greatest *in vitro*, when miR-155-5p was transfected into HCT116 cells, changes were subtle. Nevertheless, as miR-155 is highly expressed in breast cancer, this warrants further investigation as it may illustrate a relationship with aberrant methylation observed in cancerous tissue (G. Zhang *et al.*, 2015).

1.9.2 Post-translational modification of Dnmt1

Interactors may also possess enzymatic activity and chemically modify Dnmt1 by adding or removing a post-translational modification (PTM). Although Dnmt1 is reportedly decorated with PTMs (www.phospsite.org) few have been fully characterised (Jeltsch and Jurkowska, 2016). Furthermore, a recent publication from the Bestor group that is discussed below, ought to stimulate a discussion as to whether manipulation of Dnmt1 levels through overexpression can bring about false candidate PTMs (Yarychkivska, Tavana, *et al.*, 2018). What also needs to be considered is the proportion of Dnmt1 that would be modified by

endogenous levels of the interacting partner and whether the effect would manifest given there is a surplus of Dnmt1 in cells required to maintain DNA methylation (Gaudet *et al.*, 2003, 2004; Schermelleh *et al.*, 2007). Those PTMs explored so far associate with both a stimulatory and inhibitory effect on activity and protein stability. A selection of these are discussed below and summarised in Table 1.2

1.9.2.1 Phosphorylation

Adenosine monophosphate (AMP)-activated protein kinase (AMPK) mitigates the effects of dysregulated mitochondrial function by promoting the activity of PGC-1a (peroxisome proliferator-activated receptor gamma coactivator-1a), a transcription factor whose downstream targets include nuclear respiratory factor 1 and -2 (NRF1 and NRF2), which regulate mitochondrial biogenesis (Marin *et al.*, 2017).

In a HUVEC (Human Umbilical Vein Endothelial Cell) model, DNMT1 activity was shown to be impaired by AMPK-mediated phosphorylation in response to cellular stress. An interaction with phosphorylated RBBP7, a coactivator of histone acetyltransferase 1 (HAT1), then reportedly facilitated the transcription of *PGC-1a*, *NRF1* and *NRF2*, amongst other genes. This coincided with increased mitochondrial mass and citrate synthetase and electron transport chain activity. Using a line in which DNMT1 is mutated so that it cannot be subject to AMPK phosphorylation, these cellular responses did not occur. The implication from this study is that reduced Dnmt1 activity leads to a permissive state for genes that regulate mitochondrial biogenesis to be expressed (Marin *et al.*, 2017). This requires further deconvoluting as AMPK activation leads to cell cycle arrest (Marin *et al.*, 2017) and therefore dilution of the methylation patterns from impaired Dnmt1 activity would not be able to account for the hypomethylated gene promoters.

Phosphorylation of mouse Dnmt1 at serine 146 (S146) was identified and initial experiments using DNA-cellulose substrates suggested that this modification reduces the affinity of Dnmt1 for DNA. In an *in vitro* catalytic assay however, the implications of this were not detected as the phosphorylated form performed similarly to a non-phosphorylated Dnmt1 (Sugiyama *et al.*, 2010). An additional experiment to uncover an effect, if any, could be to titrate the amount of phosphorylated and non-phosphorylated protein in the *in vitro* assay as the minimal amount of Dnmt1 that may have bound in initial experiments could have been sufficient to achieve comparable rates of methylation.

An example of a phosphorylation that increases human DNMT1 activity is that at S154, and potentially the orthologous mouse S152. As this is mediated by cyclin-dependent kinases, this raises the possibility that it may occur in a cell-cycle dependent manner however, it is constitutively phosphorylated in the mouse (Lavoie and St-Pierre, 2011). Phosphorylation of residues within the N-terminal domain, like S154, have proven to induce Dnmt1 activity *in vitro* by stabilising an interaction with the C-terminal catalytic domain (Goyal *et al.*, 2007). This explanation alone is insufficient and needs to be examined in the context of the conformational changes that occur *in vivo* when present at replication foci.

Phosphorylation of the N-terminus of DNMT1 has also been suggested to regulate stability. Although western blots in the paper suggest a loose relationship, Sun *et al.* were prompted to follow the serendipitous observation that Phosphatidylinositol 3-kinase/protein kinase B (PKB) positively correlates with DNMT1 levels in multiple normal, transformed and cancer lines. An antagonistic effect of DNMT1 phosphorylation was found against ubiquitin-mediated degradation, likely mediated by a downstream effector of PKB, glycogen synthase kinase 3β (Gsk3β) (Sun *et al.*, 2007). As Gsk3β inhibition in mESCs does not affect Dnmt1 levels, the universality of this mechanism is unclear (von Meyenn *et al.*, 2016).

The influence of phosphorylated S143 (human) and the adjacent methylation of K142 have been characterised as an interdependent switch that controls DNMT1 stability. As phosphorylation by AKT1 during mid to late S phase leads to an accumulation of DNMT1 in HEK-293 cells, this antagonises SET7-mediated methylation through steric hindrance. Depletion of AKT1 in HeLa cells reduced total DNMT1 protein levels and conversely, SET7 overexpression led to reduced S143phospho species. Whilst levels of Lys142 methylation were shown to be most abundant between the end of G2 and the start of the next G1 phase, total endogenous DNMT1 levels and the modified forms subtly fluctuate when accounting for the margins of error depicted from quantified western blots. Nevertheless, DNMT1 abundance does reach a maximum when the phosphorylated species is more prevalent than the methylated form (Esteve *et al.*, 2009; Estève *et al.*, 2011). The overexpression of the mediators therefore may over emphasise the actual extent that these modifications affect endogenous DNMT1 levels *in vivo.*

1.9.2.2 Ubiquitination

Ubiquitin was referred to earlier in the context of Dnmt1 recruitment to histones. However, this stable moiety is also associated with directing protein localisation, interactions and as exemplified here, initiating proteasomal degradation (Komander and Rape, 2012).

Telomeres are specialised regions of heterochromatin that cap the end of chromosomes and offer protection from degradation. These stretches shorten with ageing in somatic cells and this results in cellular senescence. Multiple pathways are active in ESCs to ensure that telomeres are restored, thus enabling self-renewal. One such mechanism utilises zinc finger and SCAN domain-containing protein 4 (Zscan4) which is a characteristic marker, along with the retrovirus derived transcript Mu-ERV-L, of the 2-cell state that mESCs enter transiently. This state imbues mESCs with totipotent characteristics typical of early development (Morgani *et al.*, 2013). Expression of Zscan4 and Mu-ERV-L correlates with low protein levels of Dnmt1 and Uhrf1 and incidentally cells in this state are globally hypomethylated. In a Uhrf1-dependent manner, Zscan4 interacts with both Dnmt1 and Uhrf1 and instructs Uhrf1 to ubiquitinate itself and Dnmt1, likely precipitating proteasome-mediated degradation (Dan *et al.*, 2017). In mESCs lacking Dnmt1 or Dnmt3a/b, telomere lengths are significantly longer than WT counterparts. Whilst the mechanism that leads to this increase has not been fully explained, the hypomethylated cells undergo more frequent telomeric sister-chromatid exchange events, suggesting that DNA methylation is a negative regulator of this process (Gonzalo *et al.*, 2006).

1.9.2.3 SUMOylation

Small ubiquitin-related modifier (SUMO) is a moiety resembling ubiquitin that similarly acts on lysine residues. Multiple potential targets in human DNMT1 have been found, spanning lysine residues between positions 412-1616. *In vitro* assays indicate that sumoylation has the potential to accelerate DNMT1 activity and overexpression of a SUMO ligase enhances binding of DNMT1 to chromatin, but that it is not essential (Lee and Muller, 2009).

1.9.2.4 Acetylation

Bridging the N and C-terminal domains of Dnmt1 are a series of GK repeats. It has been reported that these are subject to acetylation which enables Uhrf1 to ubiquitinate DNMT1 and mark it for degradation. Deacetylation was shown to promote binding of ubiquitin-specific protease 7 (USP7) and prevent this event (Du *et al.*, 2010; Felle *et al.*, 2011; Qin,

Leonhardt and Spada, 2011). However, in a more recent study, when endogenous USP7 was reduced to near undetectable levels in MEFs and human lung carcinoma cell line H1299, there was no subsequent effect on DNMT1 levels nor DNA methylation (Yarychkivska, Tavana, *et al.*, 2018). The ramifications of this extend to other studies and should prompt a re-evaluation of conclusions reported for other PTMs in which overexpressed proteins were required to demonstrate an effect. Given the variety of different cell lines used across similar studies it would be interesting to investigate whether there are cell specific effects that may account for the discrepancies as not all the criticisms stand: knockdown of endogenous USP7 and HDAC1, that potentially removes acetylated moieties from DNMT1, did have negative effects on DNMT1 levels in LS174T and RKO colorectal cancer cells, respectively (Du *et al.*, 2010; Felle *et al.*, 2011)

Of the examples discussed so far, the PTMs have been located in the N-terminal domain yet acetylation at the catalytic portion has also been found at Lys1349 and Lys1415 (human). When these are mutated to arginine residues to act as a lysine mimic in an exogenously expressed construct, the recovered lysate exhibits greater methyltransferase activity by 2-fold *in vitro* relative to an overexpressed WT protein, suggesting that acetylation here is inhibitory. *In vivo*, deacetylation was shown to be mediated by SIRT1 however the biological context in which acetylation may be regulated was not explored (Peng *et al.*, 2011).

1.9.2.5 Poly-ADP-ribosylation

Utilising NAD, poly ADP-ribose polymerases (PARP) generate polymers of between 2-200 units that are either held in the catalytic domain or free and can bind non-covalently to proteins bearing a specific motif. One such example is DNMT1 that has two motifs in the N-terminal regulatory domain. Lysates from L929 fibroblasts subjected to PARP inhibition were more active in a methyltransferase activity assay, compared to the control, whilst free and PARP1-bound polymers were sufficient to inhibit DNMT1 activity by up to ~90%. L929 fibroblasts were further used to show an *in vivo* interaction between PARP1 and DNMT1. Whilst the authors speculate this mechanism may be important at CGIs (Reale *et al.*, 2005), it would be interesting to investigate a potential role in DNA methylation reprogramming in early development that will be discussed later (Hajkova *et al.*, 2010).

Table 1.2 – Post-translational modifications of Dnmt1 and the respective effects

Modification	**Method**	**Effect**
Phospho S143 (human) by AKT1	*In vitro*, Cos-7 cells, overexpressed with AKT1	Phosphorylation by AKT1 inhibits methylation of adjacent K142 by SET7, promoting stability (Estève *et al.*, 2011)
Phospho S146 (mouse) by CK1δ/ε	Brain lysates using GST-mDnmt1 as bait, *in vitro*	Reduces DNA binding affinity (Sugiyama *et al.*, 2010)
Phospho S152 (mouse)/S154 (human) by CDKs 1/2/5	HEK-293 cells, overexpressed, endogenous and *in vitro*	Increased Dnmt activity relative to non-phosphorylated form (*in vitro* with lysates from transfected HEK-293 cells) (Lavoie and St-Pierre, 2011)
Phospho S154 (mouse)	Murine Erythroleukemia cells, insect cells	Increased Dnmt1 activity *in vitro* compared to non-phosphorylated form (Glickman, Pavlovich and Reich, 1997; Goyal *et al.*, 2007)
Phospho S717 (mouse)/730 (human)	HUVECs, overexpressed transfection	Facilitates interaction with RBBP7; reduced activity in nuclear extracts from HUVECs and MEFs; hypomethylation of promoters in MEFs, HUVECs and *in vivo* (aorta) (Marin *et al.*, 2017)
Methylation K142 (human) by SET7	*In vitro* and HeLa cells, endogenous	Knockdown of SET7 in HeLa cells led to a corresponding increase in Dnmt1; overexpression in HeLa were 10% globally hypomethylated (Esteve *et al.*, 2009; Estève *et al.*, 2011)

SUMO – multiple along the protein	HCT116, overexpressed transfection and endogenous	Enhances Dnmt1 activity (*in vitro*) and binding (HCT116) (Lee and Muller, 2009)
Ubiquitination	mESCs, endogenous and overexpressed	Expression of Zscan4 correlates with reduced Dnmt1 protein levels, likely driven by Uhrf1 mediated ubiquitination (Dan *et al.*, 2017)
Acetylation K1111, K1113, K1115, K1117 (mouse)	HEK-293T cells, overexpressed, HDACis	Potentially involved in mediating proteasomal degradation or conferring *de novo* methylation function (Du *et al.*, 2010; Felle *et al.*, 2011; Peng *et al.*, 2011; Qin, Leonhardt and Spada, 2011; Yarychkivska, Shahabuddin, *et al.*, 2018)
Acetylation Lys1349 and Lys1415 (human)	HEK-293T cells, overexpressed, HDACis	Reduces DNMT1 activity in a catalytic activity assay (Peng *et al.*, 2011)

1.9.3 Post-translational modification of Uhrf1

The interactome of Uhrf1 is discussed in Chapters 4 and 6. Below I have outlined the influence a number of Uhrf1 PTMs confer on its activity. Of note, Uhrf1 has been annotated with fewer PTMs compared to Dnmt1 (www.phospsite.org).

Apparent from the studies summarised below is that Uhrf1 may have multiple DNA methylation independent functions. One such example discussed later entails its role in mediating the DNA damage response. Despite this, a study found that Uhrf1 protein declines with time upon irradiating cancer cells with UV light. Uhrf1 downregulation was initiated by the phosphorylation of a serine residue (human S108) situated within a degron (pSXXDpSG) by the kinase CK1δ. Consequently, this phosphorylated degron acts as a substrate for the

$SCF^{\beta\text{-}TrCP}$ E3 ligase complex which is involved in the sequential ubiquitination of proteins for proteasomal mediated degradation (Chen *et al.*, 2013) . These findings raise further questions with respect to how cells ensure there is sufficient Uhrf1 to regulate aspects of the DNA damage response, as discussed later, and what the biological relevance of downregulating Uhrf1 upon UV-induced DNA damage is. Furthermore, this also outlines a mechanism through which Uhrf1 protein stability may be regulated, a process that becomes clear is important in Chapter 3.

Uhrf1 protein levels are also subjected to Pim1 dependent regulation (Yang *et al.*, 2017). This kinase is constitutively active and can facilitate tumorigenesis by synergising with other oncogenes. However, its overexpression can also induce cellular senescence, an irreversible state of cell-cycle arrest. In this context, Uhrf1 was phosphorylated at S311 by Pim1 when the kinase was overexpressed in a fibroblast line, promoting Uhrf1 degradation. Conversely, when Uhrf1 was knocked down, markers of cellular senescence increased. These complementary observations led to the authors concluding that Uhrf1 degradation by Pim1 results in the expression of senescence associated genes, such as p16, as the genes promoters are hypomethylated (Yang *et al.*, 2017). Whether hypomethylation of the p16 promoter through Uhrf1 downregulation alone is sufficient to drive expression or unrelated cell signalling events from Pim1 overexpression are accountable for the p16 levels require further dissecting. Nevertheless, other models of senescence, namely late passage primary fibroblasts, also display aberrant DNA methylation patterns (Cruickshanks *et al.*, 2013), as discussed later. Furthermore, knockdown of Uhrf1 was sufficient to induce senescence in primary cell lines in a separate study, suggesting Uhrf1 may have a central role (Jung *et al.*, 2017).

Another phosphorylation event implicated in UHRF1 protein stability is S652 (human), mediated by CDK1-cyclin B. This UHRF1 phosphorylated variant peaked at the M-phase of the cell cycle in HCT116 cells, whilst total UHRF1 levels were lower, compared to other stages. The inverse relationship between phosphorylation and protein levels were attributed to an impaired interaction with USP7, a deubiquitinase that would otherwise protect UHRF1 from being degraded via the proteasome (Ma *et al.*, 2012). As a decline in Uhrf1 levels results in senescence in primary cells lines, as shown above, and levels may fluctuate in cancer lines during the cell cycle, there unsurprisingly appears to be an uncoupling of the influence UHRF1 may exert on the cell cycle in cancer models.

Of note, as this study manipulated levels of the endogenous USP7 and a correlation with UHRF1 stability was noted (Ma *et al.*, 2012), these findings are spared some of the criticisms levelled at similar studies into the relationship between USP7 and DNMT1, as discussed above.

The region of USP7 that mediates the interaction with Uhrf1 is a UBL domain (Ma *et al.*, 2012). Whilst Uhrf1 also bears this domain, as discussed above, it would be of interest to determine whether this directly activates Dnmt1. *In vitro* assays suggest that the UBL domain of Uhrf1 can increase the Dnmt1 activity by 2-fold (T. Li *et al.*, 2018) as addressed earlier. However, this is complicated by the evidence showing Usp7 deubiquitinates histones in *Xenopus* extracts (Yamaguchi *et al.*, 2017), which is an essential marker for Dnmt1 recruitment in mESCs (Qin *et al.*, 2015). It is therefore important to consider how these various interactions are integrated in a cell-specific manner.

The importance of modifying Uhrf1 in a developmental context has been illustrated in zebrafish. Phosphorylation at S648 (human 661) is conducted by the cyclin-dependent kinase CCNA2/CDK2. When a WT or a nonphosphorylatable *uhrf1* mRNA was coinjected with a *uhrf1* morpholino, only the WT form improved survival. Using immunofluorescence, it was shown that mutant uhrf1 is present in the nucleus of all cells whilst the WT protein resides in the nucleus and cytoplasm of some tissues (Chu *et al.*, 2012). Whilst this goes against the usual perception of the function of Uhrf1, this is particularly interesting as it has been shown to shuttle out of the nucleus in models of murine pre-implantation development, raising the possibility Uhrf1 may have a role in the cytoplasm (Mulholland *et al.*, 2018). Otherwise, the result from the zebrafish may demonstrate it is important to have temporal regulation of Uhrf1, which is abrogated with the mutant uhrf1 as it cannot shuttle into the cytoplasm.

1.10 Tets

Despite the stability of 5mC, it can be erased through a passive means wherein Dnmt1 activity is impaired and therefore the epigenetic marks are not maintained on newly synthesised daughter strands (Arand *et al.*, 2015). Alternatively, active measures, mediated by the ten-eleven-translocation (TET) enzymes, can iteratively oxidise 5mC to 5hmC, 5-formyluracil (5fC)- and 5-carboxylcytosine (5caC). The latter two products are then susceptible to excision from DNA as there is a destabilised interaction between the pyrimidine base and sugar-

phosphate group. Removal is precipitated by Thymine DNA glycosylase (TDG), a Uracil DNA glycosylase (UDG), which leaves an abasic site. This is subsequently cleaved by AP endonuclease 1(APE1), to create a single stranded break, before DNA polymerase-β, XRCC1 and DNA Lig3 coordinate ligation of the strand and restore an unmodified cytosine base (Wu and Zhang, 2017).

Isoforms of two of the TETs, TET1 and TET3, are expressed in somatic tissue but also during junctions of epigenetic reprogramming, including in the early embryo. The products of TET activity are distributed heterogeneously across the genome and they are highly influenced by chromatin state. As such, TETs are enriched at enhancer and DNAse-hypersensitive sites with low CpG density and moderate methylation. On the other hand, whilst TET1 is enriched at CGIs marked with H3K4me3 in mESCs of highly expressed genes, 5hmC levels are low, which is keeping with observations that 5mC, the initial substrate of TETs, is also usually absent. Across gene bodies, 5hmC levels increase from the transcriptional start site, at which they are depleted, to the termination site. The relationship with these regions suggests that direct recruitment by regulatory factors can direct TET activity. These distribution patterns however appear to be cell specific (Tahiliani *et al.*, 2009; Kohli and Zhang, 2013; Wu and Zhang, 2017).

Owing to its abundance in various tissues however, 5hmC should not only be regarded simply as a 5mC derivative on its way to excision, but potentially as having a function. For example, levels are 40% of all 5mC in cerebellar Purkinje neurones whilst half this amount is present in heart tissue. This abundance of 5hmC in this cell type is likely related to their postmitotic state as replication-dependent dilution does not occur. Conversely, products of 5hmC are scarce across all tissues (Kriaucionis and Heintz, 2009; Kohli and Zhang, 2013). In addition, 5hmC and indeed 5f-and 5-caC are binding site for proteins that show little affinity for 5mC, and these readers can form complexes with other chromatin modellers. Inversely, these marks can also repel 5mC readers. Whilst the number of 5hmC readers is modest in mESCs, levels are greatest in adult brain tissue for example in which 5hmC is high. This could be explained by a change in the absolute protein level of the reader in the different cell types in some cases (Spruijt *et al.*, 2013). As roles independent of catalytic activity are also evident, as well as the capacity to physically prevent Dnmt recruitment to loci, further work is required to dissect the specific function of 5hmC from the TETs (Wu and Zhang, 2017).

During PGC specification, which will not be discussed at length, and post fertilisation of the oocyte, a wave of demethylation that erodes marks on the paternal and maternal genome occurs. Evidence supports that this process entails synergistic passive and TET-mediated active pathways (von Meyenn *et al.*, 2016).

1.11 Methylation life cycle

The genome undergoes extensive epigenetic remodelling in pre-implantation development (Smallwood and Kelsey, 2012). Genome scale analyses have been employed to map the progressive loss of methylation that occurs upon fertilisation and up to the formation of the inner-cell mass (ICM) and implantation at E4.5. This epigenetic reprogramming occurs asymmetrically with the paternal complement of the genome losing marks at a relatively faster rate than the maternal component (Smith *et al.*, 2012; Peat *et al.*, 2014) (see Figure 1.9). The mechanisms that execute this epigenetic reprogramming can be broadly classified as passive or active. A passive measure to promote hypomethylation involves the impairment of methylation maintenance, meaning that DNA methylation patterns will not be restored with progressive rounds of cell division. With time, 5mC will therefore be diluted. In active demethylation for example, Tet3 can be employed to propagate the direct removal of 5mC from DNA by instigating DNA repair mechanisms.

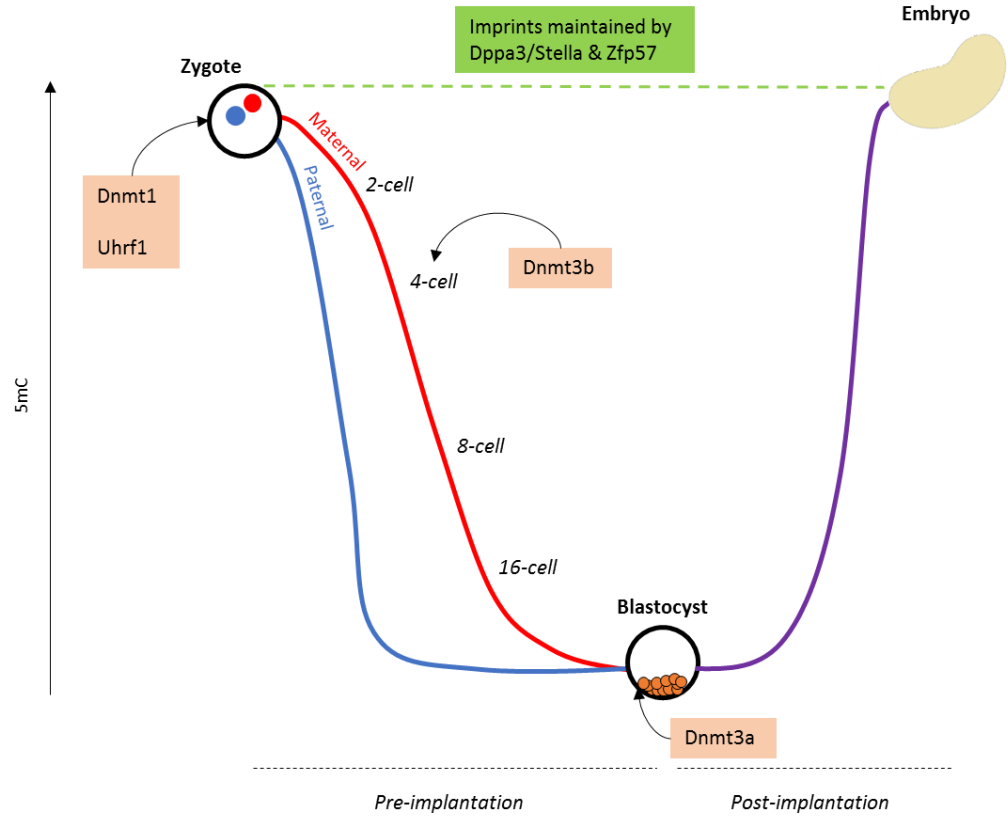


Figure 1.9 – DNA methylation is reprogrammed after fertilisation. The paternal complement of the genome (blue line) is subject to both passive and pervasive active mechanisms of demethylation. Meanwhile, the maternal genome (red line) is also reprogramed but is protected from extensive active mechanisms. Methylation of repeat sequences derived from both genomes are typically preserved throughout this window, as are marks at imprints due to the action of Stella and Zfp57. After implantation of the blastocyst around E4.5, new methylation patterns are established in a tissue specific manner under the influence of histone marks and transcription during differentiation. The time at which the *de novo* transferases, Dnmt3a and –b, are first expressed is shown. Dnmt1 and Uhrf1, an important co-factor for Dnmt1 activity, are expressed throughout albeit at relatively lower levels than differentiated cells. Adapted from Smallwood and Kelsey, 2012.

In this section, I outline the evidence supporting the notion that impairment of DNA methylation maintenance is the predominant mechanism through which demethylation occurs in pre-implantation development.

Generally, in seeking to understand the biological importance of demethylation in pre-implantation development, maintaining the view that transcription is negatively affected by

DNA methylation would be narrow and to ignore the evidence discussed above that some transcription factors have a preference for methylated motifs (Yin *et al.*, 2017). Imposing a model in which DNA methylation regulates transcription in one direction therefore needs to be replaced with a nuanced view at individual motifs and transcription factors. Subsequently the impact of inhibiting DNA demethylation on gene expression should be viewed through the prism of this specific approach.

1.11.1 Demethylation

Generally, intergenic regions and gene bodies incur the greatest loss of methylation after fertilisation, with LINEs also enriched as examples reprogrammed (Peat *et al.*, 2014; Shen *et al.*, 2014). Gene regulatory elements such as enhancers, miRNAs and long intergenic non-coding RNAs (lincRNAs) also follow the trend toward hypomethylation, typically from a higher starting methylation state in spermatocytes relative to oocytes (C. Li *et al.*, 2018). At the blastocyst stage, methylation levels of the ICM are approximately 20% of the starting combined methylation state, with trophoectodermal cells that develop into the placenta having a slightly lower amount (Maenohara *et al.*, 2017). Data collected from human equivalents at similar stages show that comparable dynamic changes also occur. Exceptions include sequence-specific repetitive elements and satellites however, this reflects the differing starting methylation level across species (Smith *et al.*, 2014; C. Li *et al.*, 2018).

Comparatively, oocytes are roughly two-fold more hypomethylated globally relative to sperm but are hypermethylated in the context of CGIs that reside in intragenic transcriptionally active regions (Smallwood *et al.*, 2011; Santos *et al.*, 2013). In this regard, oocytes have been described as having a distribution more similar to early embryonic stages wherein there is a weak correlation with local CpG density and methylation. On the other hand, post-implantation embryos, somatic tissue and sperm are highly methylated at lower CpG densities and marks tend to be excluded from CpG dense regions (Smallwood *et al.*, 2011; Smith *et al.*, 2012).

Throughout the window of epigenetic reprogramming in pre-implantation development, Dnmt1o is expressed. This is a truncated Dnmt1 isoform from an oocyte-specific transcript that lacks the first 118 N-terminal residues. Despite an estimate that it is 2000-fold more abundant than full-length somatic Dnmt1 (Dnmt1s), this isoform resides in the cytoplasm for all but the 8-cell stage when it is trafficked into the nucleus and contributes to the maintenance of imprints (Ratnam *et al.*, 2002) and X-chromosome inactivation in

extraembryonic tissues (McGraw *et al.*, 2013). Contrary to this, Dnmt1s is detected in nuclei throughout the majority of the pre-implantation sequence; a weaker association is initially observed with the paternal pronucleus but on the onset of cell division, levels are comparable to the maternal pronucleus. The two-cell stage sees Dnmt1s transit out the nucleus in G2, possibly through exportin-mediated export, otherwise it remains nuclear and chromatin associated in subsequent S phases (Kurihara *et al.*, 2008). Dnmt1s levels notably increase after the 2-cell stage, coinciding with zygotic gene activation (ZGA), and peak at the blastocyst stage (Cirio *et al.*, 2008). In parallel, faint IF signals of Uhrf1 are also detectable throughout pre-implantation development where, as in other tissues, it is important for maintaining DNA methylation. Knock-out of the maternal gene decreases methylation levels at both maternal and paternal imprints, in addition to regions that are resistant to reprogramming in control cells (Maenohara *et al.*, 2017).

The Dnmt3a *de novo* transferase meanwhile is oocyte derived yet levels beyond the 8-cell stage and up to implantation are undetectable. Conversely, Dnmt3b is distinguished as it is produced in the embryo on the onset of division to a 4-cell state and accumulates thereafter (Hirasawa *et al.*, 2008).

Low levels of the Dnmts and Uhrf1 are likely contributing factors to the progressive loss in methylation however, the extent to which requires experimental validation as mouse models have an excess pool of Dnmt1 (Gaudet *et al.*, 2003, 2004; Schermelleh *et al.*, 2007).

1.11.1.1 Active demethylation in pre-implantation development

Immunofluorescence staining illustrated that the paternal genome is subjected to replication-independent 5mC oxidation by Tet3 throughout pronuclear stages. Additional evidence from *Tet3*-null models demonstrated that this process contributes mainly at single-copy loci, such as *Oct4* and *Nanog* although others were not able to assign any genomic features that could distinguish these cases from other methylated sequences (Hashimoto *et al.*, 2012; Guo *et al.*, 2014). Whilst Tet3 becomes more diffuse in the cytoplasm at later stages, the high 5hmC signal persists up to the 8-cell stage and therefore Dnmt1 will continue to be disposed to its effects as it has an inability to restore 5mC at an opposing 5hmC mark (Hashimoto *et al.*, 2012; Arand *et al.*, 2015). At most genomic loci however, particularly those densely methylated in sperm, *Tet3*-null zygotes displayed only a minor defect in demethylation. Furthermore, clustering of methylation sequencing data of pronuclei placed *Tet3*-null models and WT closer together than patterns from null and WT pronuclei treated

with a cell-replication inhibitor. This suggested that Tet3-independent passive demethylation is the major mechanism that promotes epigenetic reprogramming in pre-implantation development (Peat *et al.*, 2014; Shen *et al.*, 2014).

The biological significance of Tet3 activity has nevertheless been investigated with conditional deletion in oocytes; when females with this mutation were mated with WT males, there was a significant reduction in viable offspring. This may be associated with the compromised demethylation and subsequent expression of *Oct4* from the 4-cell stage onward as, determined using an *EGFP-Oct4* transgene as a proxy. At the blastocyst stage however, the levels of *EGFP-Oct4* were comparable to WT mice, so this may be unable to wholly account for the lethal effect, despite it being suggested as such. Whilst pre-implantation development progresses normally in this circumstance, and transplanted embryos implant, prenatal development occurs at a reduced rate. Conversely, *Tet3* deletion from male germ cells could be compensated by WT females in heterozygous offspring (Gu *et al.*, 2011).

One study suggested that broadly, demethylation of genes is coincident with transcriptional activation, an intriguing prospect given the enrichment of ZGA examples at the two-cell stage and the effects of *Tet3* deletion on viability (Gu *et al.*, 2011; Peat *et al.*, 2014). Others have shown that the absence of *Tet3* in the mouse has no significant impact on gene expression profiles, including ZGA and the small number that are downregulated are not enriched for examples of genes directly dependent on Tet3 for 5mC removal (Shen *et al.*, 2014). In equivalent human stages of development, the global trend also suggests an uncoupling of methylation and gene expression (Smith *et al.*, 2014).

As referred to earlier, Tets can act at the beginning of a process to actively remove DNA methylation. The products that Tets generate then serve as substrates for other enzymes that can lead to the restoration of a cytosine base (Wu and Zhang, 2017). The biological importance of these enzymes have been investigated across multiple studies, as discussed below. Collectively however, the conclusions minimise the role that active mechanisms likely play in pre-implantation demethylation.

Tdg can cooperate with Ape1 and Xrcc1 to excise 5caC derived from the iterative oxidation of 5mC by Tets and 5-hydroxymethyluracil (5hmU) from the deamination of 5hmC by activation-induced deaminase (AID) activity. Conceivably Tdg may contribute to demethylation in pre-implantation development yet mRNA levels are negligible and

knockout models are embryonic lethal at E11.5. This corresponds to findings *in vitro* whereby the impact on gene regulation is minimal in ESCs and the effect is strongest in differentiated lineage committed cells (Morgan *et al.*, 2004; Cortázar *et al.*, 2011; Cortellino *et al.*, 2011; He *et al.*, 2011; Pfaffeneder *et al.*, 2014; Wu and Zhang, 2017). Tdg is also highly unlikely to act directly on 5mC substrates as *in vitro* analyses suggest it has a 30-40-fold greater activity against base-pair mismatches (Gehring, Reik and Henikoff, 2009). Levels of 5mC are also comparable between zygotes lacking maternal *Tdg* and WTs (Guo *et al.*, 2014).

An alternative mode of demethylation similarly involving AID proposes that unmodified cytosine bases are first subject to deamination, forming a uracil base, before repair is initiated by uracil glycosylase (UNG2). Long-patch base excision repair (BER) would then precipitate a loss of adjacent 5mC marks. Whilst cytosine has been shown as the preferred substrate of AID above modified forms, owing to steric hindrance imposed by methyl groups, (Hashimoto *et al.*, 2012; Santos *et al.*, 2013; Schiesser *et al.*, 2013) such a mechanism raises questions regarding the specificity and the purpose of invoking the initial cytosine deamination event. Other work has cast doubt on the extent to which the mechanisms involving AID and the related enzymes APOBEC could contribute to epigenetic reprogramming. Overexpression of these components in HEK-293T cells had no measurable effect on 5mC or 5hmC levels (Nabel *et al.*, 2012).

Nevertheless, Xrcc1, Ape1 and Parp1, another BER factor, are present in the male pronuclei of zygotes and were seen to localise to chromatin on the onset of epigenetic reprogramming. The separate chemical inhibition of Parp1 and Ape1 was sufficient to significantly reduce the demethylation of 5mC in zygotes as measured by immunostaining of paternal pronuclei. A significant effect upon LINE-1 elements only emerged with Parp1 inhibition, suggesting specificity of the mechanisms (Hajkova *et al.*, 2010). It is important to note that Parp1 can directly inhibit Dnmt1 activity, which is currently unexplored in the context of pre-implantation development (Reale *et al.*, 2005). Perversely, Parp1 can modify the SRA domain of Uhrf1 to increase its binding to Dnmt1, as revealed in co-immunoprecipitations (Co-IP) from NIH 3T3 cells. Furthermore, Parp1 reduced the apparent Uhrf1 directed ubiquitin-mediated degradation of Dnmt1, although additional controls in a repeat experiment would be necessary to definitively conclude this. Nevertheless, a *Parp1*-null cell line had equally efficient loading of Dnmt1 onto heterochromatin as a WT line and the DNA methylation

profile at these regions was also comparable (De Vos *et al.*, 2014). This suggests the effects on Dnmt1 may dominate in pre-implantation development.

A common mechanism invoked by active demethylation processes entails the recruitment of DNA damage repair enzymes. In a study into the regulation of DNA damage in the zygote, it was observed that *Xrcc1* deletion in the oocyte led to increased DNA damage markers in zygotes prior to the first round of cell division however, these lesions resolved in interphase. Furthermore, knockout of a cohesion subunit, *Scc1*, similarly led to the accumulation of DNA damage markers in the paternal pronuclei in G2 phase of the cell-cycle. Consequently, there was an extended interphase arrest or Chk1-dependent delay into mitosis in these lines. In this background, chemical inhibition of Tet3 significantly reduced the abundance of DNA damage markers and partially abrogated the block with an improvement in the number of zygotes that entered mitosis recorded. Live-cell imaging demonstrated that if left unrepaired, DNA lesions generated by Tet3 activity lead to anaphase bridges and embryo fragmentation (Ladstätter and Tachibana-Konwalski, 2016).

1.11.1.2 Passive demethylation in pre-implantation development

Notably, despite the potential contribution of Tet3, inhibiting DNA replication at the zygotic to 2-cell stage transition prevents reconfigurations of DNA methylation at most loci, illustrating a dependency on passive modes of demethylation. This is supported by the persistence of hemi-methylated CpG dyads throughout pre-implantation development, reflecting impaired methylation maintenance (Guo *et al.*, 2014; Shen *et al.*, 2014; Arand *et al.*, 2015).

As a representative genomic class, LINEs exhibit the most pronounced demethylation during the sperm to zygotic stage in humans and mice (Smith *et al.*, 2014). In the mouse, ~50% of LINE-1 CpG dyads are on average present in a hemimethylated state, albeit distinct families are more resistant. In contrast, demethylation at ERV and SINEs progresses relatively uniformly (Shen *et al.*, 2014; Arand *et al.*, 2015). Such a loss may be causative for their transcriptional activity, shown to be important for pre-implantation development due to effects on chromatin conformation and gene transcription (Kigami *et al.*, 2003; Beraldi *et al.*, 2006; Thompson, Macfarlan and Lorincz, 2016; Jachowicz *et al.*, 2017). It has been suggested however that their demethylation may be counteracted by *de novo* methylation after the 2-cell stage, despite the rapid downregulation of the machinery at equivalent stages, before a sharp drop is again promoted prior to the blastocyst stage (Arand *et al.*, 2015). Of note, the

absence of *Tet3* has no significant impact on the expression of these elements either (Inoue, Matoba and Zhang, 2012; Shen *et al.*, 2014).

In contrast to the male complement, the maternal genome undergoes some hydroxymethylation at a broad range of genomic elements but is largely protected from pervasive Tet3 activity; the extent of which however is disputed (Guo *et al.*, 2014; Peat *et al.*, 2014). Nevertheless, a small overlap in regions actively demethylated in both male and female pronuclei during early pre-implantation development have been identified (Guo *et al.*, 2014). The differential rate of demethylation has been attributed to the action of the maternal effect gene Dppa3/Stella and its binding partner RanBP5. The absence of *Stella*, or a perturbation in its nuclear localisation, promotes extensive demethylation of the maternal genome in pronuclear stages, impacting some imprinted genes and repeats (Nakamura *et al.*, 2007). Its mode of action is dependent on the G9a deposited H3K9me2 mark, to which it binds with high affinity. As this mark is particularly enriched across the maternal pronuclei the rate of demethylation proceeds more slowly, possibly through an inhibitory effect on Tet3 and BER components (Hajkova *et al.*, 2010; Nakamura *et al.*, 2012).

1.11.2 Inheriting DNA methylation patterns

Heritable methylation patterns, defined as such as they deviate from the global trends and are resistant to loss throughout cleavage divisions, have been identified by comparing regions hypermethylated in either gamete and hypomethylated in the other. The quantity of regions contributed by sperm are greatest and reside in intergenic regions, in contrast to oocytes which are CGI-associated. These loci are highly methylated in the respective gamete and are distinct from imprinted genes, which retain the sex-specific pattern against the global loss of methylation (Smallwood *et al.*, 2011). With respect to reconfigurations in the human, methylation dynamics are represented by 10 varying patterns. Notably, there is local maintenance of patterns across gene bodies; nearly two thirds were shown in one study to retain at least some methylation at the blastocyst stage compared to sperm. Divergent regulation is evident at methylated regions derived from oocytes, but this contrasts with ICRs which are conserved. Diversity between mouse and human extends to the demethylation and transcription of repeat elements, owing to species specificity (Smith *et al.*, 2014).

Zebrafish inherit parental methylation patterns and are not subject to global erasure in embryogenesis. Paternal patterns are identical to those of pluripotent embryos whilst maternal marks are gradually erased concurrently with cleavage divisions against

a background of histone pattern remodelling (Murphy *et al.*, 2018). Throughout the growth of the pre-implanted embryo and two phases of ZGA, an increasing number of promoters become more accessible, as determined by ATAC-seq, and therefore the corresponding genes are more likely to be transcribed. Whilst there is some correlation between the time at which a specific site is more accessible and CpG density, the strongest predictor is the extent of DNA methylation; more densely methylated regions are associated with delayed accessibility. This is likely to be driven by CXXC domain-containing factors that bind preferentially to unmethylated CpG sequences. This relationship however doesn't hold true when applied to distal regulatory regions that similarly emerge during ZGA. These have relatively high levels of DNA methylation and are enriched for binding sites for the zebrafish Oct4 homolog and Nanog, which induce a local hypomethylation when engaged. Cell signalling changes will therefore likely dictate the binding of these factors and not simply the removal of DNA methylation (Liu *et al.*, 2018). If there is evidence of the association with residual methylation and accessibility in the mouse, this may enable a function to be assigned to these marks that are not associated with imprinted regions.

Mosaic hypomethylation patterns associated with ZFP57 mutations provided evidence that this Kruppel-associated box-containing zinc-finger protein (KRAB-ZFP) protects the methylation status of ICRs genes post-zygotically, and the establishment of a subset in oocytes (Li *et al.*, 2008; Mackay *et al.*, 2008). This function occurs through the binding to a methylated hexanucletoide sequence situated within an ICR, prior to recruitment of KRAB-associated protein 1 (KAP1), a scaffold for a number of histone modifying proteins. Immunoprecipitations in ESCs have also demonstrated a direct interaction with Dnmts -1, -3a, -3b and Uhrf1, thus potentially creating an environment that can withstand the global loss of methylation in pre-implantation development (Quenneville *et al.*, 2011). Indeed, sites targeted by ZFP57 are methylated in oocytes and these marks are retained as such in ESCs and pre-implanted blastocysts (Strogantsev *et al.*, 2015).

Inactivation of *Zfp57* is embryonic lethal by E14.5–E16.5 whilst the use of mESCs models shows an effect on the expression of only 9 imprinted genes. The aberrant expression is generally associated with the acquisition of H3K4me3 at the Zfp57 binding site, which would further impede the reestablishment of methylation patterns. Interestingly, there is an enrichment of H3K4me1/3 at these sites in differentiated tissues; this mark may be present on the unmethylated allele otherwise this would suggest that the effects of

Zfp57 supersede the inhibitory effect of H3K4 methylation in dictating the activity of *de novo* transferases (Riso *et al.*, 2016). Zfp57 binding sites are interesting loci to study for how histone methylation marks influence the deposition of DNA methylation patterns. In TKO cells for example, there is a loss of H3K9me3 at Zfp57 binding sites, supporting a model in which Zfp57 binding is dependent on DNA methylation, enabling recruitment of histone modifying complexes (Quenneville *et al.*, 2011). However, in *Zfp57*-null mESCs, there is a loss of both H3K9me3 and DNA methylation (Riso *et al.*, 2016), suggesting that Zfp57-driven histone methylation operates within a feedback loop.

1.11.3 Implantation and remethylation

The establishment of new methylation patterns at low and intermediate density CpG promoters is coincident with uterine implantation, predominantly through the action of Dnmt3b. On average, methylation is a poor predictor of gene expression but correlates strongly with silencing when present within a CGI. This is exemplified by germ-line specific and X-linked genes and some tissue-specific transcription factors (Weber *et al.*, 2007; Borgel *et al.*, 2010; Isagawa *et al.*, 2011). In most cases, the underlying methylation pattern is dictated by neighbouring histone marks, which are also subject to renewal during early development (Mikkelsen *et al.*, 2007; Meissner *et al.*, 2008; Isagawa *et al.*, 2011).

The majority of DNA methylation is deposited in the E6.5 epiblast with promoters of pluripotency related factors such as *Tcl1* being subject to modification. Lineage specific differences between early germ layers are limited but methylation accumulates as development progresses. A further set of pluripotency genes, represented by *Rex1* and *Stella,* become modified at E9.5, and tissue specific methylation patterns are adopted (Borgel *et al.*, 2010; Isagawa *et al.*, 2011). The differentiation of multipotent stem cells is accompanied by small changes in promoter methylation and the majority are relayed to daughter cells (Weber *et al.*, 2007; Sørensen *et al.*, 2010). Marks deposited across clusters of intermediate CpG density promoters are also potentially erased in a tissue specific manner during terminal differentiation, exemplified by brain, eye and haematopoietic specific genes in the respective tissues (Borgel *et al.*, 2010).

ESCs are counterparts to the ICM of E3.5 blastocysts from which the embryo derives. As this population retains their developmental potential when explanted, as evidenced by their ability to contribute to the three germ layers in chimeras, their enforced differentiation *in vitro* has proven a useful tool in tracking the dynamics of methylation as specific cell fates

are adopted (Marks and Stunnenberg, 2014). Despite the development of alternative culture methods that will improve the veracity of the pluripotent state (discussed later), they have provided further evidence for the accumulation of DNA methylation and the influencing factor of histone marks; high density CpG regions in neural progenitor cells that lose H3K4me3 partially acquire 5mC and this is accentuated when there is a complete loss of H3K27 and H3K4 methylation at previously bivalent promoters (Meissner *et al.*, 2008). Separately, it was shown that promoting the differentiation of mESCs through application of retinoic acid stimulates G9a to target DNA methylation to the *Oct3/4* locus, whereby the modification and heterochromatin formation occurs after the silencing of the gene and it demonstrably 'locks' in a transcriptionally repressive state (Feldman *et al.*, 2006; Tachibana *et al.*, 2008).

This relationship has also been corroborated in hESCs stimulated to differentiate randomly. Gene ontology analysis inferred that gains in methylation occur predominantly at genes related to developmental processes, such as organogenesis and anatomical structure development. In this study however, the CGIs that behaved in this way were notable for their enrichment at the 3' end of genes. Furthermore, the tissue specific differential methylation correlated positively with expression. An overlap of this stretch of nucleotides with a CTCF binding motif suggests a mechanism analogous to that which is operative at the *H19*/*Igf2* imprint control region. Here, methylation would abolish CTCF binding and prevent the insulating effect it would have on the ability of an enhancer to act on an upstream promoter (Yu *et al.*, 2013).

The tracking of hESCs induced to differentiate into motor neurones illustrated differential dynamics of the *de novo* DNMTs during lineage fate commitment: DNMT3A levels were sustained highly throughout whereas DNMT3B was rapidly downregulated upon initiation. As expected, genes and regulatory elements associated with alternative lineages acquire methylation in differentiating WT cells but more prominent is the demethylation at genes associated with spinal cord and neural tube development. In the absence of DNMT3A, global methylation levels in progenitor and differentiated cells are comparable to those from a WT line however, differentiation is significantly impaired, as is the viability and fitness of differentiated cells. Within the window of differentiation, intermediate progenitor cells from the *DNMT3A*-null line were shown to express markers associated with the gliogenic lineage more highly than a neuronal lineage. Cross referencing the genes with aberrant methylation

and expression returned a list of 148 DMRs and these tended to overlap with gene enhancer elements. An absence of methylation at a regulatory element of *PAX6*, a master neurogenic factor, failed to promote its expression as this likely permitted the binding of a repressor.

Conversely, *DNMT3B* knockout has no significant effect during this specific differentiation pathway on efficiency, despite the number of DMRs that arise compared to WT cells being much greater than in the case of *DNMT3A*-null hESC differentiation (Smith *et al.*, 2014; Ziller *et al.*, 2018). This further presses the need to unpick relationships between methylation and transcriptional changes, which ultimately drive the fates of differentiating cells.

In addition to promoters, methylation patterns are established at the enhancers of pluripotency genes following differentiation cues, primarily due to Dnmt3a recruitment in response to H3K4me1 and H3K27ac erasure by the Lsd1-Mi2/NuRD complex. This activity correlates with reduced expression of the associated gene (Petell *et al.*, 2016). These distal regulatory elements are regions at which most of the changes in DNA methylation actually occur during differentiation (Meissner, 2010). This transition is also in part influenced by the binding of pioneer factors that can protect remodelling at tissue specific loci. For example, FoxD3 binds to the Alb1 enhancer during endoderm differentiation, which proceeds diversification to a hepatocyte lineage, and Ptcra enhancers are protected by Sp1 and E protein in thymocyte specification (Xu *et al.*, 2009). This aforementioned study however was concerned only with the correlation between DNA methylation and transcription factor binding, leaving the biological relevance of the relationship to be explored. It nevertheless puts into context the influence transcription factors exert over DNA methylation machinery.

The relationship between gene expression, specifically that of variant native transcripts, and the methylation of CpG patches distal to promoters is displayed too at the *Zdbf2* locus. This gene is expressed in multiple regions in the mouse brain. Following fertilisation however, it remains methylated on the maternal allele and is regarded as being transiently imprinted throughout pre-implantation development. During this developmental window, a longer isoform is expressed briefly from a transcription site residing within a paternal germline DMR. Upon implantation and acquisition of methylation across the locus on both alleles, H3K27me3 is evicted and this induces the derepression of a shorter form expressed from a somatic DMR upstream of the promoter (Greenberg *et al.*, 2017). This mechanistic study elegantly encapsulates the interplay between epigenetics, both methylation and histone

modifications, and transcription, whilst also exploring the functional role of partially demethylated loci.

1.12 Methylation in disease and altered cell states

Methylation patterns are not necessarily maintained indefinitely in somatic tissues, with changes observed in ageing, cellular senescence and more dramatically in cancer and ICF.

1.12.1 Senescence and ageing

Senescence describes the process of cell proliferation arrest, which can prevent tumour formation. A panel of morphological and gene expression changes are associated with this cell state, in addition to modifications to chromatin and an altered DNA methylation profile. In this case, there is global hypomethylation, predominantly at late replicating and laminin associated domains that are partially methylated in the equivalent primary cell line and at lowly expressed genes. Modest hypermethylation is also evident at early replicating domains and is significant at the periphery of genes repressed in senescence, such as cell cycle regulators. The mechanism that imposes the differential effects would be interesting to uncover, with suggestions including a redistribution of Dnmt1. This becomes more pertinent as the methylation patterns in senescent cells, with respect domains and genes predictive of clinical outcome, resemble those in cancer (Cruickshanks *et al.*, 2013).

The functional significance of the DNA methylation changes in senescence are currently unknown. This is similar to the examples seen with ageing, in the so-called epigenetic clock. Variable levels of DNA methylation can arise at CGIs, bivalent domains and Polycomb targets with age. Interestingly, these changes can be used to develop models that predict biological age from the methylome of a wide variety of tissues. Changes have been shown to occur more rapidly during earlier years however, this is unlikely to reflect mitotic age as non-proliferative tissues also exhibit differences. Furthermore, cancerous tissue tends to deviate from this model and in the case of breast cancers, different stratified tumour classes display variable rates of ageing (Horvath, 2013). This particular observation supports the need to consider transcriptional changes, and the inputs that influence these, when deciphering the mechanism through which the alterations in DNA methylation arise.

1.12.2 Cancer

A comprehensive review of the role of DNA methylation in cancer is beyond the scope of this thesis however, comments on some important findings are found below.

The observation that aberrant DNA methylation is present in nearly all human neoplasms makes it an attractive source of investigation in the pursuit of therapeutic agents, particularly in cases of hypermethylated tumour suppressor genes (Jones and Baylin, 2002). Indeed, epigenetic medicines, such as DNMT inhibitors (DNMTi), have FDA approval for the treatment of acute myeloid leukaemia (AML) and their use in combinatorial therapies with other drugs targeting epigenetic modifiers has been explored (Jones, Issa and Baylin, 2016).

A role in gene silencing and the potential for spontaneous deamination, resulting in a C to T transition, are possible mechanisms through which DNA hypermethylation could promote cancer development. Studies in colon cancer cell lines suggest that signalling pathways which regulate Dnmt activity are also perturbed: impairing DNMT1 had little effect on global methylation levels and it required combinatorial knockouts of *DNMT1* and *DNMT3B* to confer a notable change. Furthermore in the same study, silenced *p16* was re-expressed and the growth rate of the line was significantly reduced upon application of a demethylating agent (Jones and Baylin, 2002).

This case is an illustration of the 50 % of genes shown to be mutated in familial cancer that also undergo hypermethylation. Other silenced genes of predicted importance for cancer development have no consistent mutation and therefore the main driver of their suppression is potentially epigenetic (Jones and Baylin, 2002).

Experimental models have also shown a causal link between DNA hypomethylation and neoplasia. Compound heterozygous mice carrying null and hypomorphic *Dnmt1* alleles were generated and the expression of the protein was 10% of WT levels. The majority of mice developed thymic tumours, with nearly all tumours exhibiting a gain of multiple chromosomes and c-myc overexpression. A follow up study determined that hypomethylation can promote loss of heterozygosity (Eden *et al.*, 2003; Gaudet *et al.*, 2003). Although these models are extreme cases of hypomethylation, they serve as a proof of principle that perturbation of DNA methylation levels can be causative. Nevertheless, large numbers of malignant tumours and metastases with low methylated cytosine content have been screened. An increase in LINE-1 and ERV mobility, and subsequent mutational

insertions, expression of proliferation associated genes and chromosomal rearrangements from demethylated satellite sequences were noted (Ehrlich, 2002).

Although the examples above provide compelling evidence for a role of dysregulated DNA methylation in cancer, to establish that hypermethylation is neoplastic, it is important to evaluate the expression of aberrantly methylated genes in both cancerous and preneoplastic tissue. For this, the specific cell type from which the cancer arose within a tissue, the presence of alternative promoters and whether the methylation is a secondary event following an upstream silencing event are important considerations (Bestor, 2003; Sproul and Meehan, 2013), as are evaluations of copy-number alterations (Martin-Trujillo *et al.*, 2017). There is also evidence that promoters of CGIs associated with genes found frequently hypermethylated in a large variety of cancers are typically silent in the tissue of origin prior to neoplasia. The majority of hypermethylated genes may therefore be regarded as passenger genes, distinct from the driver genes whose methylation can promote neoplasia (Sproul *et al.*, 2011, 2012). The importance of assessing transcriptional profiles is further underscored by observations that *in vitro* culture of primary cell lines can impose changes to the methylation profile (Varley *et al.*, 2013; Nestor *et al.*, 2015).

Furthermore, the therapeutic benefits of DNA demethylating agents may not necessarily depend on the expression of silenced tumour suppressor genes but rather the promotion of an immune response driven by the presence of dsRNAs from endogenous repeats that become expressed. This causes a cascade of antiproliferative and pro-apoptotic interferon response genes that can engage the immune system (Chiappinelli *et al.*, 2015; Roulois *et al.*, 2015)

1.12.3 ICF

A disease that brings to light secondary structural roles of DNA methylation is ICF (immunodeficiency, centromere instability and facial anomalies). This is characterised in humans by hypomethylated satellite 2 and 3 of constitutive pericentromeric heterochromatin of chromosomes 1, 9 and 16 in all cell types. Thread-like chromosomes prone to breakage are also found in lymphocytes (Xu *et al.*, 1999; Jin *et al.*, 2008). Patients can be stratified depending on the additional demethylation of alpha satellites and females can also present a hypomethylated inactive X-chromosome however, perturbations in expression of X-linked genes are minimal. Sequencing has shown that the majority of cases can be explained by point mutations within the catalytic domain of Dnmt3B, resulting in

substitutions or splicing defects that impair the protein function. In compound heterozygous patients, nonsense mutations within the N-terminus of the gene can also be causative (Jiang *et al.*, 2005).

Furthermore, mutations that cause substitutions, splice defects and interestingly, given the absolute requirement in mouse models, null alleles of the interacting partner Lsh/HELLS are associated with the emergence of the disease (Thijssen *et al.*, 2015). Deleterious mutations in the zinc finger domains of cell division cycle associated 7 (CDCA7) and *ZBTB24*, which encodes a transcription factor important for regulating CDCA7, and more generally B- and T-cell differentiation, account for other cases (de Greef *et al.*, 2011; Thijssen *et al.*, 2015). The mechanism in which these mutations cause ICF has recently come to light with use of a *Xenopus* model. Under these circumstances, there is reduction in Lsh loading at chromatin due to a disrupted interaction with CDCA7. Together, these act in a bipartite complex to remodel chromatin, which is demonstrably a prerequisite for DNA methylation at satellite sequences (Jenness *et al.*, 2018).

The phenotype of ICF provides evidence for the role of DNA methylation in maintaining chromatin integrity and for the specificity of the methyltransferases as evidently, DNMT3A or DNMT1 could not compensate for reduced DNMT3B function. It would be of interest to determine whether the recruitment of DNMT3B to the hypomethylated sequences still occurs in the disease and if this occludes DNMT3A from compensating. Further work is also required to dissect heterogeneous transcriptional changes in patients. Evidence suggests that altered promoter methylation is not a sufficient explanation for all genes and may be an effect of redistributed H3K27me3 (de Greef *et al.*, 2011).

1.12.4 Rett Syndrome

A rare disease related to the function of DNA methylation is Rett syndrome. Mechanistically, the Rett syndrome phenotype is attributed to the reduced binding of mutated MeCP2 to methylated DNA and with other complexes that influence transcription, potentially through histone modification. Downstream, many cell specific effects have been observed, encompassing a variety of biological processes (Leonard, Cobb and Downs, 2017).

Patients have reduced brain volume and poor synaptic connectivity between neurones. Conditional knockout mouse models yield a phenotype only in the region from which it was deleted, and additional consequences emerge with whole knockouts. This includes fatty liver

and aberrant bone structure, suggesting a role of MeCP2 in non-neuronal tissue (Leonard, Cobb and Downs, 2017). The importance of this factor extends beyond early development as post-natal deletion in adult mice results in Rett syndrome-like symptoms (McGraw, Samaco and Zoghbi, 2011). Conversely, rescue reverses many of the defects (Guy *et al.*, 2007).

The overwhelming majority of patients are female and because the gene is X-linked, it is typically fatal in males. However, mutations in MeCP2 that do not manifest as Rett syndrome in females can lead to mental retardation in males (Chahil, Yelam and Bollu, 2018).

The molecular mechanism of this disease illustrates the complex relationship between DNA methylation and the histone code and the cascade of events that develop under particular circumstances in different cell types.

1.13 Modelling pluripotency

The culture of mESCs in conventional medium, supplemented with either FCS or BMP4 and the cytokine LIF, is appropriate to maintain a heterogeneous population of differentiated and pluripotent cells (Ying *et al.*, 2003). A contributing factor to this heterogeneity is the capacity of LIF to engage the gp130 receptor, which simultaneously leads to activation of the pluripotency promoting STAT3 transcription factor and the pro-differentiation Ras-Erk signalling cascade. Whilst cells integrate the insulating effects of serum or BMP4 that regulate levels of Inhibitor of Differentiation (Id) proteins via R-Smad protein phosphorylation (Ying *et al.*, 2003; Marks and Stunnenberg, 2014), the networks are insufficient to wholly prevent differential expression of markers associated with ESCs and PGCs. This is best represented by variable levels of *Stella, Nanog* and *Oct4* across the population of cells cultured in this medium. In addition, cells representative of an intermediate state between ESCs and late epiblast stem cells (EpiSCs), which markedly express *Fgf5,* can also co-exist. Separating the populations by these markers and testing the respective developmental potential demonstrates that the high *Nanog* and *Oct4* expressing cells are more efficient at generating embryoid bodies and are resistant to experimentally induced differentiation (Hayashi *et al.*, 2008).

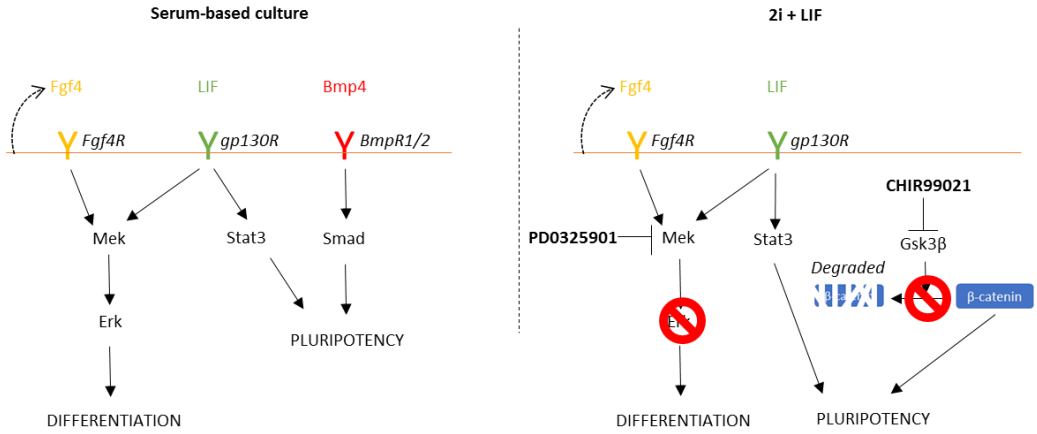


Figure 1.10 - Serum-based culture is insufficient to shield the pro-differentiation effects of Fgf4 which is secreted from mESCs in an autocrine manner. Additionally, LIF promotes the maintenance of pluripotency however, as it also influences Mek activity, it has antagonistic effects. Media supplemented with serum or one particular component of this, Bmp4, activates Smad signalling and this positively regulates Id proteins that stabilise the undifferentiated state. In 2i+LIF media, Mek activity is abrogated with the addition of PD0325901. Furthermore, the degradation of β-catenin is prevented through the inhibition of Gsk3β with CHIR99021. This enables β-catenin to disrupt Tcf3, which would otherwise negatively affect the expression of pluripotency genes. Adapted from Marks and Stunnenberg, 2014

The development of an alternative medium referred to as 2i + LIF has enabled mESCs to be harvested from pre-implanted blastocysts and propagated *in vitro* to an extent that ground state pluripotency is more closely recapitulated (Ying *et al.*, 2008; Marks *et al.*, 2012). This term refers to the developmental period before inductive fibroblast growth factor 4 (Fgf4) signalling promotes specification of the ICM into the hypoblast, the source of extraembryonic tissue such as the yolk sac, and the priming of the naive epiblast to respond to developmental cues and differentiate into germ layers that constitute the embryo (Silva and Smith, 2008; Nichols and Smith, 2009).

Prohibiting the differentiation of mESCs in culture driven by mitogen-activation protein kinase (Mek) and the canonical Wnt pathway component Gsk3β, is achieved with the inhibitors PD0325901 and CHIRON99021 (CHIR99021), respectively. Collectively, when used

in a chemically defined base medium with LIF, this is termed 2i+LIF (Ying *et al.*, 2008). From hereafter, it will be referred to as 2i.

The rationale for developing a model system in which Mek is inhibited was explained by the evidence that Fgf4 exerts an effect through this signalling axis. Furthermore, exposure of pre-implanted blastocysts to 2i prevents the emergence of the hypoblast and the ICM becomes entirely pluripotent epiblast (Nichols *et al.*, 2009; Wray, Kalkan and Smith, 2010). As Fgf4 is produced in an autocrine manner by undifferentiated ESCs, this could prompt a commitment to differentiate into various lineages whereas in the presence of PD0325901, this influence is abrogated. This is a direct result of reduced Erk1/2 phosphorylation, an intermediate in Mek signalling that leads to neuronal or mesodermal differentiation (Kunath *et al.*, 2007).

At the molecular level, Gsk3β inhibition results in the stabilisation of β-catenin and this interacts directly with Tcf3 and alleviates its binding to chromatin. This Tcf/Lef transcription factor would otherwise negatively influence the expression of core pluripotency transcription factor targets (see Figure 1.10) (Silva and Smith, 2008; Wray, Kalkan and Smith, 2010; Wray *et al.*, 2011).

The majority of transcriptional changes occur within 48 hours upon transitioning mESCs from serum-based cultured to 2i media (Hackett *et al.*, 2013; Galonska *et al.*, 2015). Whilst ~75% of genes are expressed at similar levels, that cells are maintained more robustly in a pluripotent state in 2i is reflected in reduced expression of ectoderm and mesoderm specification genes relative to serum media. There is also higher expression of *Prdm14* which can act to supress extraembryonic endoderm transcription, but markers of this lineage are generally expressed to a similar extent (Marks *et al.*, 2012; Marks and Stunnenberg, 2014).

On the other hand, a greater number of transcripts associated with stem cell maintenance are present at higher levels within serum culture. Nevertheless, in 2i there is a uniform expression of the pluripotency regulators *Rex1* and *Nanog*, which may further antagonise signalling mediated through Mek (Ying *et al.*, 2008). Furthermore, *Nanog* is biallelically expressed in 2i, thus recapitulating the transient state *in vivo* when this occurs in early development. The expression of other important auxiliary genes expressed in serum culture are absent in 2i, such as *c/N-Myc*, which underscores the dramatic transcriptional rewiring that occurs (Marks and Stunnenberg, 2014).

Proteomic analysis suggests that many of the transcriptional changes have a corresponding effect on protein abundance. In accordance with gene expression data, Gene Ontology and KEGG analyses stipulate that proteins involved in metabolic processes, linked to the reduction in c/N-Myc levels and the influence of Mek and Gsk3β on glucose processing, are more enriched in 2i. Meanwhile, mesoderm and ectoderm development classes are represented strongly in serum culture (Ying *et al.*, 2008; Taleahmad *et al.*, 2016).

Various iterations of the base media to which 2i is added impose differential outcomes in mESCs on the expression of genes and retrotransposons, particularly those of the ERV class, that are associated with either primed or naïve pluripotent states. Distinct clusters are evident in principal component analyses between conditions with 2i additives and those without, which reflect more naïve and primed states of pluripotency respectively. Application of either inhibitor individually promotes the expression of master primary germ layer genes but these are nearly extinguished in combination, demonstrating the interdependency to dampen the effects of differentiation promoting networks. Another distinction is the expression of genes associated with the two-cell stage of pre-implantation development, such as Zscan4. These are more highly expressed in a 2i base media containing KnockOut™ Serum Replacement (KOSR) compared to N2B27 (Hackett *et al.*, 2017). Further details about these two different base media are discussed in chapter 3.

Interestingly, culture of E2.5 embryos in N2B27+2i for a period of 3 days was insufficient to supress the emergence of a subpopulation of cells coexpressing both pluripotent markers and the trophoblast marker *Hex*, amongst other examples. The trophoblast is the outer epithelial layer of cells of the blastocyst that develop into the placenta. This property reflected a pre-implantation stage during which early cells are totipotent. Indeed, when *Hex* positive cells were sorted and used to form chimeras, there was evidence that they contributed to embryonic and extraembryonic lineages (Morgani *et al.*, 2013). Whether 2i therefore can sustain the propagation of a cell that has the potency of both ground state epiblast and EpiSCs would be an interesting path to follow (Brons *et al.*, 2007). Nevertheless, this observation provides further support for the use of 2i as a means to recapitulate early development and suggests that further manipulations of the media may improve the developmental potency of the cultured cells and the veracity with which individual developmental stages can be modelled.

The 2i formulation has enabled the derivation of embryonic germ cells and ESCs from multiple different mouse strains (Hanna *et al.*, 2009; Kiyonari *et al.*, 2010; Leitch *et al.*, 2010) and from the rat, which were previously intractable (Buehr *et al.*, 2008; Leitch *et al.*, 2010). The use of 2i therefore facilitates research into the fundamental mechanisms that govern the pluripotent state across cell types and species.

1.13.1 DNA methylation and 2i

Changes in the histone modification landscape occur upon switching from serum to 2i media. The most striking is the reduction in H3K27me3 levels at the promoters of lowly expressed genes and within a 5kb region up- and downstream and redistribution across satellite DNA. As the effect on H3K4me3 is slight, approximately two thirds of the 3000 bivalently marked gene promoters in ESCs resolve to being modified by H3K4me3 alone, which has little effect on the already low expression levels of the associated genes (Marks and Stunnenberg, 2014; Guo *et al.*, 2016). With respect to H3K9me2, multiple studies have noted reduced levels through ChIP and western blot methods however, there are conflicting reports as to whether H3K9me3 is similarly affected (Leitch *et al.*, 2013; Marks and Stunnenberg, 2014; Sim *et al.*, 2017).

Furthermore, global DNA methylation levels are reduced between 2-3-fold within 14 days of culture in 2i, from ~4% to ~1% of all cytosines bearing the mark. This establishes a population of cells that in this regard closely resembles the E3.5 stage of pre-implanted epiblasts. In contrast, cells maintained in serum have global levels comparable to that of EpiSCs. Methylation patterns at ~69% of gene promoters are equivalent between serum cultured mESCs and EpiSCs, compared to only 38% in pre-implanted blastocysts (Borgel *et al.*, 2010; Ficz *et al.*, 2013; Habibi *et al.*, 2013; Leitch *et al.*, 2013). Of the various genomic regions associated with gene regulation, CpG and non-CpGI promoters, enhancers and bivalent promoters are impacted, although the rate at which methylation is lost varies. Meanwhile GO term analysis suggests a lack of correlation between specific processes and methylation retention (Hackett *et al.*, 2013; Walter *et al.*, 2016). Whilst minor satellites and LINEs and SINEs are subject to methylation loss, levels at ERVs, particularly those belonging to the IAP class, major satellites and ICRs are largely maintained, and this correlates with H3K9me3 enrichment. Expectedly, as 5mC is the precursor to 5hmC, the dynamics are closely mirrored. The alterations in epigenetic patterns are also fully reversible upon switching between serum-based media or 2i. Ultimately, the resulting methylation patterns in 2i resemble that

of the early pre-implantation blastocysts (Ficz *et al.*, 2013; Habibi *et al.*, 2013; Hackett *et al.*, 2013; Leitch *et al.*, 2013).

The high 5hmC levels observed at the enhancers and promoters of some pluripotency genes in serum-based culture have been suggested to assist in maintaining a more permissive state for their expression (Hackett *et al.*, 2013). However, as an equivalent highly methylated yet pluripotent state is not present *in vivo*, the observations are likely to be an artefact of tissue culture and not the result of a coordinated mechanism.

Long-term culture of male derived mESCs in 2i can compromise their developmental potential. Methylation deposited at ICRs is eroded and this results in the aberrant expression of imprinted genes. At other DMRs that distinguish male and female ESCs and in the distribution of the histone variant H2A.X, which is situated proximal to genes that regulate early development, male derived cells can resemble less developmentally potent female ESCs with long-term 2i culture (Choi *et al.*, 2017). This is supported by findings that *Gsk3α* and *-β* double knockout lines exhibit reduced methylation at imprints and aberrant expression (Meredith *et al.*, 2015).

Demethylation in 2i media coincides with reduced *Dnmt3a*, *Dnmt3b* and *Dnmt3l* mRNA and protein levels. This is at least partially regulated by Prdm14, an important effector of PGC specification that also assists in maintaining ESCs in a pluripotent state. Levels of *Prdm14* increase in 2i and it likely exerts an effect through a dual mechanism, acting as a transcriptional repressor upstream of the respective gene promoters and initiating ubiquitination of the proteins in a complex with G9a (Grabole *et al.*, 2013; Sim *et al.*, 2017). Although downregulation of the *de novo* pathway components is a contributing factor, as they assist Dnmt1 in maintenance methylation, the rate at which hypomethylation is achieved in *Dnmt3a* and *–b*-null lines is significantly delayed compared to 2i culture (Lei *et al.*, 1996; Liang *et al.*, 2002; Chen *et al.*, 2003; Jackson *et al.*, 2004). Similarly, the increase in Tet activity in 2i was shown in knockout lines to partially contribute but not be wholly responsible for driving hypomethylation. An exception to when Tets contribute robustly is with vitamin C supplementation, a co-factor of α-ketoglutarate-dependent dioxygenases, such as Tets, that can induce their activity (Blaschke *et al.*, 2013; Hackett *et al.*, 2013; von Meyenn *et al.*, 2016; Sim *et al.*, 2017).

Evidence therefore points towards the impairment of the DNA methylation maintenance machinery as the means by which mESCs become hypomethylated in 2i. Indeed, it was shown

that Uhrf1 protein levels decline upon switching cells from serum to 2i media. A correlation was also noted between the levels of H3K9me2 at genomic loci and the extent of DNA demethylation incurred. This relationship therefore contributes to the locus specific maintenance of DNA methylation at certain genomic regions despite a global loss of the mark (von Meyenn *et al.*, 2016).

In light of these dramatic changes, one question which arises is to what extent does differential transcription between the different media depend on the epigenetic changes, such as DNA hypomethylation? Furthermore, can these conclusions be extrapolated to understand the biological relevance of DNA demethylation in pre-implantation development?

1.13.2 DNA methylation and pluripotency

The rate of demethylation varies for individual genes in 2i culture (Walter *et al.*, 2016) and expression changes are uncoupled from promoter demethylation (Ficz *et al.*, 2013). Nearly half the genes repressed and around a quarter of those induced in 2i already display a significant change within 24 hours. The three components, Oct4, Sox2 and Nanog (OSN), that constitute the core pluripotency network are subject to binding rearrangements within this same period, yet the sites at which this occurs are distinct from the regions at which H3K27me3 and H3K4me3 marks change and that exhibit DNA demethylation (Galonska *et al.*, 2015).

The differential binding of OSN nevertheless appears to influence distal regulatory elements and therefore potentially transcription. These sites are however depleted of the aforementioned histone marks and are only subjected to a gradual loss in methylation beyond 7 days in 2i culture (Galonska *et al.*, 2015). Taken together, this suggests that the outcomes relating to OSN binding at this stage are influenced directly by alterations in transcription and independent of epigenetic changes.

Whilst DNA demethylation is globally uncoupled from transcription, this does not exclude locus specific effects and it may require the use of single cell sequencing technologies to assay DNA methylation and transcription to uncover a possible relationship. One single-cell DNA methylation method is in its infancy however, it has indicated the heterogeneity of serum culture, with respect to DNA methylation, is not eliminated in 2i. This will cofound the

ability to ascertain a relationship between changes in DNA methylation and cell signalling (Smallwood *et al.*, 2014).

Distilling the approach to uncover how DNA methylation regulates transcription and pluripotency can be achieved by comparing WT and TKO cells. However, TKOs are rather a blunt tool and there is a caveat that long-term culture of this line may have permitted compensatory mechanisms to be acquired that silence aberrantly expressed genes (Fouse *et al.*, 2008). Nevertheless, 337 genes are over 2-fold upregulated in TKO cells, representing X-linked and tissue-specific genes and transcription factors, which is in line with other studies (Fouse *et al.*, 2008; Karimi *et al.*, 2011; Hackett *et al.*, 2012). Of the genes bound by Oct4 and Nanog in WT cells, which are themselves expressed at normal levels in TKO cells and are typically unmethylated in WT mESCs (Imamura *et al.*, 2006; Fouse *et al.*, 2008), ~75% of methylated genes were expressed compared to ~90% of unmethylated genes, suggesting DNA methylation has little influence in directing the activity of these factors (Fouse *et al.*, 2008). This is further supported by the observations in 2i discussed earlier (Galonska *et al.*, 2015) and by the fact that less than 2% of genes upregulated in TKO cells are bound by Oct4/Nanog. (Fouse *et al.*, 2008). As introduced above, some transcription factors have an affinity for methylation and this was shown to be the case for Oct4 (Yin *et al.*, 2017). Therefore, assessing the direct effects of DNA methylation on Oct4 binding would likely be unnecessary.

An alternative technique, RNA-FISH, was employed to track *Nanog* expression itself. This showed that DNA methylation has a subtle effect on its heterogeneous expression in serum cultured cells and signalling changes imposed by 2i were a greater influence in attaining homogeneity (Singer *et al.*, 2014).

Collectively, these studies suggest that DNA methylation has little influence in dictating the binding of the core pluripotency network. Therefore, the transcriptional profile of 2i cultured cells and the fact that it more closely recapitulates the ICM is independent of DNA methylation changes and can be accounted for largely by alternate cell signalling effects imposed by 2i. However, to what extent is the presence of any amount of DNA methylation important for pre-implantation development?

This question has been addressed by generating embryos by nuclear transfer using TKO ESCs as the donor. In this case, they developed grossly normally throughout the pre-implantation stage. Chimeric embryos showed evidence that the TKO cells contribute to the ICM at E4.5,

before residing specifically in the placenta and yolk sac at E17.5 and E18.5. Notably, embryoid bodies constituted from TKO ESCs upregulated primitive ectoderm and early mesoderm markers to a relatively higher extent than WT embryoid bodies, whilst also downregulating *Nanog* (Sakaue *et al.*, 2010). Development of the trophoblast is also initiated prior to changes in methylation patterns that define this specific layer and the pluripotent ICM (Nakanishi *et al.*, 2012).

Furthermore, mouse knockout models for the *Dnmt* genes are lethal after the event of implantation (Li, Bestor and Jaenisch, 1992; Okano *et al.*, 1999). Collectively, the above evidence argues against a role of DNA methylation in pre-implantation development.

The alternative question is whether DNA demethylation is important for the development of the pluripotent ICM. Notwithstanding the *Tet3* knockout models referred to earlier, this would be difficult to test directly as it would require maintaining high DNA methylation levels *in vivo*. Studying the dynamics throughout reprogramming of somatic cells to induced-pluripotent cells (iPSCs) may offer some indications however, the main distinction to consider is that the generation of iPSCs is an artificial system.

Upon overexpression of the well-described *OSKM* (*Oct4, Sox2, Klf4* and *c-Myc*) transcription factors in MEFs, a delayed redistribution of DNA methylation was noted to occur after transcriptional changes and histone modification occupancy. Acquisition of DNA methylation accounted for the silencing of a subset of bivalent or monovalent genes that loose either the repressive H3K27me3 or active H3K4me3 marks. Conversely, of the few protein coding genes and lncRNAs that acquire H3K4me3 and are expressed in the latter stages of reprogramming, the density of H3K4me3 correlated inversely with DNA methylation (Hussein *et al.*, 2014). However, DNA methylation dynamics were not unidirectional with domains shown to either acquire methylation gradually or during the latter stages of reprogramming, acquire or lose marks before reverting to the state prior to the transduction of *OSKM* or alternatively, to lose methylation late in reprogramming and remain demethylated in the pluripotent state. Examples of the latter include binding sites for OSK and Nanog. The extent to which this reflects reprogramming of the zygote would be an interesting comparison to establish (Lee *et al.*, 2014).

The use of Dnmt inhibitors, Dnmt1 knockdown (Mikkelsen *et al.*, 2008) and transduction of *Tet1* (Gao *et al.*, 2013) have been used to improve reprogramming efficiency. Knockdown of *Tet1*, which is consistently upregulated throughout reprogramming of MEFs, reduces colony

formation whereas overexpression has a positive effect. Against an initial background of increasing *de novo* methyltransferase expression and DNA methylation, Oct4 demethylation and expression is promoted by Tet1 in conjunction with SKM in a secondary reprogramming system. This cocktail had comparable efficiency to the transduction of *OSKM* in generating pluripotent colonies. Overall, genes derepressed in the pluripotent state had reduced 5mC at promoters, which includes *Oct4*, *Nanog* and *Rex1*. At the end point of reprogramming, most silenced genes were not associated with acquisition of 5mC, although such a correlation was observed in intermediate stages (Gao *et al.*, 2013). As *Tet*KO models die perinatally and Oct4 expression is delayed upon *Tet3* deletion, this would suggest Tets in the reprogramming system have a greater influence compared to an *in vivo* context (Gu *et al.*, 2011; Wu and Zhang, 2017).

Transposing the conclusions from reprogramming systems to *in vivo* development should therefore be done with caution. Analyses into the 2i system would also argue for a locus and time-specific approach to assess the interplay between transcription and DNA methylation. An important point to consider however, is that a pluripotent cell is only regarded as such if it has the potential for multilineage differentiation, under the appropriate culture conditions (Martello and Smith, 2014). Therefore, by definition, DNA methylation is important for pluripotency as *Dnmt*-null models are lethal, as discussed above (Li, Bestor and Jaenisch, 1992; Okano *et al.*, 1999). The use of TKO cells have already shown it is dispensable for self-renewal, another important characteristic of pluripotency (Tsumura *et al.*, 2006). Where the window of opportunity lies to further dissect the role of DNA methylation is in considering whether it is important for it to be erased in pre-implantation development.

It is therefore clear that further investigations are necessary to identify mechanisms that contribute to DNA methylation in pre-implantation development. Once these have been characterised in functional studies, attention can then turn to perturbing them *in vivo* in order to elucidate the importance of DNA methylation dynamics in early development.

1.14 Project Aims

The maintenance of mESCs in naïve and primed pluripotent states can be modelled using 2i and serum-based culture media and this is accompanied by profound changes in the global methylation profile of the cells, as discussed earlier. When this project began, there was a deficit of reports on the molecular mechanisms responsible for promoting hypomethylation in 2i media. Furthermore, whilst N2B27+2i media had been shown to effectively demethylate mESCs, an alternative system referred to as KOSR+2i lacked any similar report.

Considering the gaps in the knowledge, one of the aims of this thesis was to:

1) Identify the interacting partners of DNMT1 and Uhrf1 in primed and naïve mESCs. Proteins that differentially interact with the DNA methylation maintenance machinery between serum and 2i media would be considered as candidate regulators of DNA methylation.

The effects of PTMs that may modulate DNA methylation in the context of 2i media were also considered however, the focus was on a specific example termed protein citrullination. This aim was developed from preliminary data generated by Maria Christophorou that suggested both Dnmt1 and Uhrf1 were subjected to this modification. This was interesting in the context of 2i media as previous studies, discussed later in the relevant chapter, suggested citrullination is promoted in mESCs with its application. Therefore, a second aim was to:

2) Understand the implications of citrullination on DNA methylation in both serum and 2i media and address the underlying molecular mechanisms.

Collectively, these aims could be summarised as attempts to define the molecular mechanisms that regulate the differential levels of DNA methylation in mESCs in response to a change in the signalling environment. In the context of serum and 2i media, which model a more primed and naïve state of pluripotency in *vitro* respectively, this will begin to link specific cell signalling cascades to an underlying change in the chromatin state, which itself may have subsequent effects on cell biology. However, this information also has the potential to inform future studies investigating the biological significance of DNA methylation reprogramming that occurs *in vivo* throughout pre-implantation development.

Chapter 2 – Materials and Methods

2. 1 Cell culture

All cell lines used throughout this project were mESCs maintained on tissue culture grade flasks and plates (Corning) with 0.1% (w/v) porcine gelatin (Sigma) at 37°C with 5% CO_2. Cells were detached using Accutase™ (BioLegend®) and passaged every other day or when confluency reached 70-80%. Cells from an E14 background were cultured in Glasgow's MEM (GMEM; Gibco™) base media supplemented with 10% (v/v) FCS (GE Healthcare Hyclone), 1X Non Essential Amino Acids (NEAAs; Gibco™), 1mM Sodium Pyruvate (Gibco™), 2mM L-Glutamine (in house technical services preparation), 0.1mM 2-Mercaptoethanol (Gibco™) and 1.0 x 10^3 U mL^{-1} of LIF (Millipore). For J1 background mESC lines, including TKO-121 lines (Tsumura *et al.*, 2006), the same recipe was followed as above except Dulbecco's Modified Eagle Medium (DMEM; Gibco™) base media, in place of GMEM, and a 15% (v/v) FCS supplement were used.

Two different 2i base media compositions were used during the project. The most predominantly used and referred to as KOSR elsewhere consisted of: GMEM supplemented with 10% (v/v) KOSR (Gibco™), 1% (v/v) FCS, 1X NEAAs, 1mM Sodium Pyruvate, 2mM L-Glutamine, 0.1mM 2-Mercaptoethanol and 1.0 x 10^3 U mL^{-1} of LIF. Media referred to widely as N2B27 was used on one occasion during the characterisation of the 2i systems, as shown in Chapter 3. This was composed of a 1:1 ratio of Neurobasal™ Medium (Gibco™) and DMEM/F-12 (Gibco™) to which: 0.5X N-2 supplement (Gibco™), 0.5X B-27™ Supplement (Gibco™), 2mM of L-Glutamine, 7.5% (v/v) bovine serum albumin (BSA; Gibco™), 11.9M monothioglycerol (Sigma) and 1.0 x 10^3 U mL^{-1} of LIF were added.

From both 2i base medias, aliquots were taken and 1µM of PD0325901 (Selleckchem) and 3µM of CHIR99021 (Sigma) prepared in DMSO were added and kept for no longer than 48 hours.

Where the pan-PADI inhibitor Cl-amidine (Millipore) was utilised, it was dissolved in water and used at 200µM.

2.2 Microbiology methods

All media and LB-Agar plates for bacterial culture were made in house by technical services staff.

2.2.1 Transformations

To propagate plasmids, an appropriate strain of bacteria was transformed. Owing to the presence of the *ccd*B lethality gene, the pB Destination vector used in the Gateway® cloning introduced below was propagated with XL1 Blue Competent *E.coli* Cells (Agilent). These contain the F' episome to negate the effects of the aforementioned lethality gene. For the sequential steps throughout the Gateway® cloning of the pB BirA* construct, Library Efficiency™ DH5α™ Competent *E.Coli* Cells (Invitrogen™) were utilised. For the In-Fusion mediated cloning of constructs into the pB BirA* vector, Stellar Competent *E.coli* Cells (Takara Bio) were used. In all instances, an appropriate amount of plasmid or reaction mixture, as dictated by the manufacturer's instructions, was initially incubated with 50µL of cells on ice for 30 mins. These were subsequently heat-shocked for 45 secs at 42^{o}C, recovered on ice for 2 mins and then added to 450µL of room-temperature SOC media. After a 1 hour recovery period at 37^{o}C, various dilutions were plated separately onto LB-Agar plates and incubated overnight at 37^{o}C.

2.2.2 Minipreps

In order to screen for 'positive' bacterial clones containing the desired plasmid, individual clones were picked and grown overnight in 5mL L-Broth at 37^{o}C with shaking and appropriate antibiotics for selection. Plasmids were subsequently extracted using QIAprep Spin Miniprep Kit (Qiagen) without deviations from the protocol supplied. Diagnostic tests were carried out with appropriate restriction enzymes (New England Biolabs) in accordance with the manufacturer's instructions. The digested plasmid was then resolved on TBE-agarose gels and compared against a predicted digestion pattern generated by the program SnapGene®. Sequencing of plasmids was performed on a 3730 Series Genetic Analyzer (Applied Biosystems®) by technical services staff.

2.2.3 Maxipreps

Once the sequence of a plasmid was confirmed, larger quantities were generated for further manipulation or transfection into mammalian cell lines. To this end, the Qiagen Plasmid Maxi Kit was used without deviations from the manufacturer's instructions.

2.3 Generation of BirA* stable cell lines

2.3.1 Gateway cloning of BirA into a piggyBac vector*

Vectors encoding *BirA*,* pcDNA3.1 mycBioID*,* was a gift from Kyle Roux (Addgene plasmid # 36047 ; http://n2t.net/addgene:36047 ; RRID:Addgene_36047) (Roux *et al.*, 2012). This was initially unsuitable for expression in mESCs as the vector was optimised for viral delivery and would be susceptible to silencing. Therefore Gateway® cloning (Thermo-Fisher) was employed to transpose *BirA** into a pDestination piggyBac vector with a CMV early enhancer/chicken β actin (CAG) promoter. pPB-PGK-destination was a gift from Austin Smith (Addgene plasmid # 60436 ; http://n2t.net/addgene:60436 ; RRID:Addgene_60436) (Takashima *et al.*, 2014) . In order to subsequently clone a gene of interest and ensure it would be fused with *BirA** at the N- or C-terminus, primers were designed so that a unique restriction site not present in the piggyBac backbone was inserted also. The protocol to establish these was provided by Thermo-Fisher and the primers were designed utilising tools on Snapgene® software.

2.3.1.1 Gateway cloning - BP recombination

In Gateway cloning, the BP reaction is dependent on the use of an Insert with flanking *att*B sites that are complementary to *att*P sites in a pDONR vector. Following recombination of these sites, an Entry clone is established in which the gene of interest is flanked by *att*L sites, a substrate for the subsequent LR reaction.

To generate a *BirA** insert with *att*B sites, 1ng of the original BirA* vector was added to a mix of 12.5µL CloneAmp™ HiFi PCR Premix (Takara Bio), 0.3µM each of forward and reverse primers (Sigma-Aldrich; see Table 2.1) and a PCR reaction was performed consisting of 30 cycles of: 98^{o}C for 10 secs, 55^{o}C for 15 secs and 72^{o}C for 1 min. The product was resolved on a TBE-agarose gel and the appropriate band excised and recovered using NucleoSpin Gel and PCR Clean-Up kit (Takara Bio).

Using an online service provided by New England Biosciences (http://nebiocalculator.neb.com/#!/dsdnaamt), 50 fmol of the BirA* insert and pDONR 221 vector were calculated and the appropriate amounts incubated together with 2µL of BP Clonase II enzyme at 25^{o}C for 1 hour, in a total reaction volume of 10µL. This reaction was stopped by adding 1µL of Proteinase K at 37^{o}C for 10 mins after which 1µL was taken and transformed as described above. Colonies were screened using diagnostic restriction enzyme

digestion (New England BioLabs) and candidates were sequenced accordingly to identify the desired Entry clone.

2.3.1.2 Gateway cloning - LR recombination

Following the reaction to generate an Entry clone, the plasmid for which the sequencing data was most complete when aligned to the expected sequence was used in the LR reaction. Here, *att*L sites that flank the gene of interest in the Entry clone are recombined with *att*R sites in a pDestination vector. Accordingly, 50ng of the Entry clone was incubated with 150ng of the pDestination vector in a total volume of 8µL, to which 2µL of LR Clonase II was added. This was incubated for 3 hours at 25^{o}C and then stopped with 1µL Proteinase K and incubation for a further 10 mins at 37^{o}C. Transformation of 1µL of this mixture then followed as described above and colonies were picked, screened and sequenced.

*2.3.2 In-Fusion cloning of hDNMT1 and mUhrf1 into pB BirA**

In-Fusion cloning reagents (Takara Bio) were used to separately clone in *human DNMT1* (*hDNMT1*) and *mouse Uhrf1* (*mUhrf1*) into the assembled pB BirA* vectors, exploiting a unique restriction site so as to generate fusion proteins that are tagged both at the N- and C-terminus with BirA*. Owing to the high molecular weight of both DNMT1 and Uhrf1 and the various domains present in each protein, this approach offered an opportunity to detect a greater range of interactors. pcDNA3/Myc-DNMT1 was a gift from Arthur Riggs (Addgene plasmid # 36939 ; http://n2t.net/addgene:36939 ; RRID:Addgene_36939) (Li *et al.*, 2006) and m*Uhrf1* in a vector was kindly supplied by Tomek Jurkowski.

Using the In-Fusion cloning tool on SnapGene®, primers were designed in accordance with the N-terminal vector being cut with Pac1 and the C-terminal vector with SmaI. As PacI cuts with an overhang, to ensure that genes were in frame after ligation, the sequence 'AA' was added to the forward primer between the sequence that would anneal to m*Uhrf1* and h*DNMT1* in the template vectors, and the backbone of the pB BirA* vector.

To generate inserts, CloneAmp™ HiFi PCR Premix and primers (see Table 2.1) were incubated with 1ng of a template vector containing either *hDNMT1* or *mUhrf1* as detailed above. The program was adjusted appropriately depending on the length of the gene being amplified, in accordance with the protocol. The reaction was resolved on TBE-agarose gels and the

appropriate band excised and recovered using the NucleoSpin® Gel and PCR Clean-up kit. A no-template control was also prepared and resolved.

The pB BirA* vectors were linearised by incubating with either PacI or SmaI restriction enzymes overnight. Using the phenol-chloroform extraction method outlined below, the vector was recovered and 50ng was incubated with an amount of insert that fulfils the ratio of 2:1, insert to vector. This was calculated on http://nebiocalculator.neb.com/#!/ligation which utilises the formula: Insert mass = desired insert/vector molar ratio x mass of vector x ratio of insert to vector lengths. The total volume of this reaction was made up to 8µL and 2µL of In-Fusion HD Enzyme Premix (Takara-Bio) was added. This reaction was incubated for 15 mins at 50^{o}C, from which 2.5µL was taken and transformed into Stellar competent *E.coli* cells (Takara-Bio) as outlined earlier and colonies were picked, screened and sequenced. Transformation of the linearised vectors alone served as negative controls and gave an indication of the reaction efficiency.

2.3.3 Purifying vectors – Phenol/chloroform extraction

Following the overnight linearisation of the vectors as outlined above, 1X volume of phenol solution (Sigma) was added to each reaction mixture before vortexing briefly and centrifuging at 17,000 x *g* for 2 mins. The top aqueous layer was recovered and treated with 1X volume of chloroform (Fisher Chemical), vortexed and centrifuged at 17,000 x *g* for 2 mins. The top layer was then recovered and one-tenth volume of 3M NaOAc and 3.5X volumes of 100% (v/v) EtOH were added to precipitate the plasmid. Following centrifugation at 17,000 x *g* for 10 mins, the supernatant was removed and the pellet was washed in 1X volume of 70% (v/v) EtOH with centrifugation at 17,000 x *g* for 2 mins. The supernatant was removed and the plasmid air-dried before being suspended in nuclease-free water. The concentration was measured on the Nanondrop Nucleic Acid channel.

2.3.4 Transfections

Stably expressing cell lines were generated using transfection with Lipofectamine 2000 (Thermo-Fisher), following the protocol supplied. The quantity of DNA in each transfection was 2µg of the pBase transposase vector and 1µg of the piggyBac vector containing the gene of interest. Cells were maintained in 50mg/mL hygromycin B for up to 10 days for the selection of transfected colonies.

Other stably expressing cells generated by others, namely the PADI4 and control cells, were also established using this method.

Table 2.1 – Primers for cloning BirA* constructs

Primer	Forward	Reverse
Gateway cloning		
Cloning of *BirA** for N-terminal tagging	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATGGAACAAAAACTCATCTCA	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAATTAAGCGTACGAGGCCT
Cloning of *BirA** for C-terminal tagging	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCCCCGGGAAGGACAACACCGTGCCCCTG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGAGATCCTCTTCTGAGAT
In-Fusion® cloning		
Cloning of h*DNMT1* into N-terminal *BirA** vector	GGCCTCGTACGCTTATTATGGAGCAGAAGCTGATCTCAGAG	GAAAGCTGGGTCTTAATTAACTAGTCCTTAGCAGCTTCCTCCT
Cloning of h*DNMT1* into C-terminal *BirA** vector	GCAGGCTTCACCCCCATGGAGCAGAAGCTGATCTCAG	GGTGTTGTCCTTCCCGGGCTCCTTAGCAGCTTCCTCCT
Cloning of m*Uhrf1* into N-terminal *BirA** vector	GGCCTCGTACGCTTATTATGTGGATCCAGGTTCGAACTATGG	GAAAGCTGGGTCTTAATTAATCACCGGCCGCTGC
Cloning of m*Uhrf1* into C-terminal *BirA** vector	GCAGGCTTCACCCCCATGTGGATCCAGGTTCGAACTATG	GGTGTTGTCCTTCCCGGGCCGGCCGCTGCCATAGCCAGG

2.4 BioID experimental procedure

Culture media used throughout is as detailed above. All solutions for lysis and preparation of the beads contained protease inhibitors (Roche). For each cell line and treatment, 2.0 x 10^6 cells were seeded in triplicate on gelatin coated 10cm dishes. After ~24 hours, fresh media was applied with a supplement of 50µM biotin (Sigma-Aldrich) prepared in DMSO. Following a further ~24 hours of culture, the media was drained and the cells were washed 3 X in 10mL

ice-cold PBS on ice, whereby the PBS was left on the dish for 30 secs during each wash to ensure diffusion of free biotin out of the cells. Cells were next lysed in 1mL RIPA buffer and scraped into a pre-chilled 1.5mL tube and pulse-vortexed for 5 secs every 10 mins over a period of 30 mins; 100 U of the nuclease reagent Benzonase (Millipore) and 1mM $MgCl_2$ was added 15 mins into this period. This time point was selected as the reagent protocol suggests the concentration of SDS in the RIPA buffer would ablate Benzonase enzymatic activity after 15 mins. Lysates were centrifuged at 17,000 x *g* for 5 mins at 4°C and 22.5µL of the supernatant was taken as input and frozen immediately on dry ice. The remainder was added to 10µL of streptavidin-conjugated agarose beads (GE-Healthcare) in a new 1.5mL tube that had been prepared by thorough washing 3 X with RIPA buffer. The remaining cell pellet collected after centrifugation was frozen on dry ice. The pull-down proceeded with end-to-end rotation at 4°C for 1 hour. Beads were pelleted by centrifugation at 6,200 x *g* for 40 secs and 22.5µL of supernatant was taken as flow-through and frozen immediately on dry-ice; the remainder was discarded. The beads were then washed in accordance with the protocol for similar proximity ligation methods as follows (Branon *et al.*, 2018), using 1mL of the reagents at each step that were prepared in dH_2O unless stated otherwise: 1 X RIPA; 1 X 1M KCl; 1 X 0.1M Na_2CO_3; 1 X 2M Urea (in 10mM Tris-HCl, pH 8.0) and 2 X in RIPA. A further 3 X washes in TBS were conducted prior to submission of the samples for mass spectrometry analysis, ensuring transfer to a new tube with each wash to eliminate detergent carry-over. For the purposes of western blotting, the beads were boiled for 5 mins in 2X Laemmli-SDS buffer (Sigma) and the supernatant was taken. The input and flow-through were similarly prepared with 4X Laemmli-SDS buffer (Sigma) and boiled for 5 mins.

Prior to optimising the aforementioned method, cells were lysed in RIPA and by a 60 mins pulse of sonication (10 secs on/10 secs off) in a Bioruptor® (Diagenode). Whilst chromatin was completely solubilised, mass spectrometry analysis suggested this method caused proteins to precipitate during the pull-down as few proteins were enriched over the control. Therefore, the compromise method outlined above was developed in which a small chromatin pellet remains after clarifying the lysate however, the majority of the BirA* fusion proteins are solubilised (Figures 3.11 and 12) and no precipitate forms. An example of the data obtained from the earlier lysis method is shown below in Table 2.2. Note the small number of proteins that are enriched over the background.

Table 2.2 – A portion of BioID data obtained from cell lines prepared using an older method. This displays all interacting candidates from an experiment using the DNMT1-BirA* N-terminally tagged line that are enriched over a control (Ratio N-Dnmt1/Empty vector).

Gene names	Average Empty (LFQ)	Average N-Dnmt1 Serum (LFQ)	Ratio N-Dnmt1/Empty vector
TEX2	1	40818001	40818001
TUBAL3	1	27558001	27558001
GFPT1	1	19291667.67	19291667.67
SKA2	1	8150334.333	8150334.333
DYNC1LI1	1	7645301	7645301
DAP3	1	4960634.333	4960634.333
NAA40	1	1152501	1152501
DNMT1	550343334.3	1.69E+11	266.0593137
NUP98	947693334.3	700576667.7	0.748426038
TXN	141863334.3	90293001	0.615410308
DNAJC2	1073343334	626646667.7	0.594407961

2.5 Myc pull-down experimental procedure

The details of the culture media used are described above. Buffers for lysis and washes contained protease inhibitors (Roche). Each cell line for both serum and KOSR + 2i treatments were plated at a density of 2.0 x 10^6 cells in triplicate on gelatin coated 10cm dishes. After ~ 24 hours, media was replaced with fresh media. The following day, cells were washed on ice with 10mL ice-cold PBS and lysed with 200µL lysis buffer containing: 10mM Tris-HCl; 150mM NaCl; 0.5mM EDTA; 0.5% NP-40 and 5 U of Benzonase were added with 1mM $MgCl_2$. The tube was left on ice for 30 mins and pulse-vortexed at 10 min intervals before centrifugation at 17,000 x *g* for 10 mins at 4°C. The supernatant was transferred to a pre-chilled tube containing 300µL of the dilution/wash buffer: 10mM Tris-HCl; 150mM NaCl and 0.5mM EDTA, from which 22.5µL was taken as input. The remainder was transferred to a pre-chilled tube containing magnetic beads conjugated to a MYC-trap (Chromotek) that had been washed 3 X in the dilution/wash buffer. This was end-to-end rotated for 1 hour at 4°C. The beads were then magnetically separated from the lysate, of which 22.5µL was taken as flow-through and the remainder discarded. For 3 X, the beads were washed in ice-cold wash buffer

and prepared appropriately as above for mass spectrometry analysis or western blot, along with the remaining cell pellet, input and flow-through.

2.6 BioID and Myc pull-down data analysis

Data for the BioID screen was initially processed by calculating the mean LFQ intensity for each individual protein identified from triplicate samples. The average LFQ value for experimental samples were then divided over the corresponding value in the Empty vector control and separately, the free BirA* control. A two-tailed unpaired t-test was also applied to filter out proteins that had significantly (p>0.05) variable LFQ intensities between replicates in an experimental condition, compared to a control condition. Proteins that were at least two-fold and significantly (p≤0.05) enriched in an experimental line over the Empty or free BirA* control were taken forward for consideration.

The Myc pull-down data was similarly analysed by dividing the average LFQ intensity of a protein in an experimental line over the average intensity in the pull-down control in the same culture conditions. Proteins enriched at least 2-fold were then filtered further using a t-test to identify those significantly enriched.

The approach taken to identify proteins from the BioID and Myc pull-down screens that differentially interact with DNMT1 or Uhrf1 in either serum or KOSR+2i is discussed in Chapter 4. Volcano plots with imputed values were generated by Jimi Wills (IGMM, University of Edinburgh) in R using a script available at the following address: https://github.com/wasimaftab/LIMMA-pipeline-proteomics/blob/master/limma_main.R.

2.7 Co-immunoprecipitation experimental procedure

The culture and lysis of cells is as described in the above procedure: 'Myc pull-down experimental procedure'. Following the centrifugation step to clarify the lysate, the supernatant was recovered and diluted with 300µL of the dilution/wash buffer: 150mM NaCl and 0.5mM EDTA. From this, 22.5µL was taken as input and the remainder was incubated overnight at 4^{o}C with rotation in a new microfuge tube with 10µL of Protein G Magnetic Beads (Thermo-Fisher); these had been prepared earlier by washing 3 X in the dilution/wash

buffer. The following day, 2µg of antibody (see Table 2.3) was added and left rotating at 4°C for a further 2 hours. Next the beads were magnetically separated from the

lysate and 22.5µL was taken as flow-through and the remainder discarded. The beads were then subjected to 3 X washes in ice-cold wash buffer with agitation.

2.8 Western blotting

2.8.1 Protein lysate preparation for western blotting

Lysates were prepared from cells washed 1 X in ice-cold PBS with modified Laemmli buffer (120mM Tris-HCl pH 8.0; 4% (w/v) SDS; 20% (v/v) glycerol). After boiling for 5 mins, chromatin was sheared with a 15G needle and syringe until the consistency was no longer viscous. Protein was quantified on the Nanodrop using the Protein A280 channel and aliquots taken and normalised with Laemmli and a 10X loading buffer (β-mercaptoethanol, 1% (v/v) bromophenol blue), before boiling for a further 5 mins. Samples were either loaded immediately into gels or stored at -80°C until required.

2.8.2 Experimental procedure

Samples were resolved on precast 4-20% Tris-Glycine gradient polyacrylamide gels (Bio-Rad) at 200V with a Tris-Glycine SDS Running buffer (25mM Tris-HCl, 190mM glycine, 0.1% (w/v) SDS). Proteins were subsequently transferred onto 0.2µm pore-sized nitrocellulose membranes (Bio-Rad) at 800mA for 1 hour using a Tris-Glycine Transfer buffer (25mM Tris-Cl, 190mM glycine, 20% (v/v) methanol). Membranes were next blocked in 5% (w/v) BSA in TBS-T (Tris-buffered saline, 0.1% (v/v) TWEEN-20) at room temperature for 1 hour with gentle agitation. Primary antibodies were then applied in blocking buffer at appropriate dilutions (see Table 2.3), typically overnight at 4°C with agitation. The following day, membranes were washed with TBS-T over 45 mins, wherein the buffer was changed every 15 mins, before the blots were incubated in appropriate horseradish peroxidase (HRP) conjugated secondary antibodies (see Table 2.3) in blocking buffer for 1 hour at room temperature with agitation. After washing membranes over another 45 min period, bands were visualised on the chemiluminescence channel on an ImageQuant LAS visualiser.

Table 2.3 – Antibodies for western blotting and immunofluorescence

Reactivity	Clone	Dilution	Host
Western blotting – Primary antibodies			
Apobec3	B-2	1:200	Mouse
Aust2	(ab96326)	1:500	Rabbit
Dazl	3/11A	1:500	Mouse
Dnmt1	K18	1:200	Goat
Dnmt3a	64B1446	1:200	Mouse
Dnmt3b	52A1018	1:200	Mouse
Dppa3/Stella	B-2	1:200	Mouse
Gapdh	GA1R	1:5000	Mouse
Histone H3	mAbcam 10799	1:2000	Mouse
Histone H3 (citrulline R2)	EPR17703	1:2000	Rabbit
Myc	9B11	1:5000	Mouse
Padi2	(ab50257)	1:1000	Rabbit
Padi4	EPR20706	1:1000	Rabbit
Pdk1	4A11F5	1:200	Mouse
Streptavidin (HRP conjugated)	-	1:2000	-
Uhrf1	M-132	1:2000	Rabbit
Utf1	(ab24273)	1:1000	Rabbit
Western blotting – Secondary antibodies (HRP conjugated)			
α-Goat	-	1:5000	Rabbit
α-Mouse	-	1:5000	Rabbit
α-Rabbit	-	1:5000	Goat
Immunofluorescence – Primary antibodies			
Dnmt1	K18	1:1000	Goat
Uhrf1	D6G8E	1:250	Rabbit
Myc	71D10	1:200	Rabbit
Immunofluorescence – Secondary antibodies			
α -Goat- Alexa Fluor® 488	-	1:200	Donkey
α -Rabbit- Alexa Fluor® 555	-	1:200	Donkey

2.9 Quantitative polymerase chain reaction

2.9.1 Isolation of RNA

RNA was harvested from cells in accordance with the Qiagen RNeasy Plus Mini handbook. RNA was eluted with RNAse-free water and stored at -80^{o}C until required.

2.9.2 RNA quantification and cDNA synthesis

Eluted RNA was quantified on the Nanodrop channel A260. Equal concentrations of each sample in an experiment were taken for further processing into cDNA; this was typically between 200ng-1µg. To this, 250ng of random primers (Promega) and 10mM dNTPs (Life Technologies) were added and the volume was made up to 13µL with RNase free water. After incubating at 65^{o}C for 5 mins, 1µL of SuperScript™ III Reverse Transcriptase was added, along with the requisite 1µL DTT, 4µL First Strand Synthesis Buffer (Promega) and 20U of RNasin® (Promega). The reaction then proceeded at 25^{o}C for 5 mins, 50^{o}C for 60 mins and 70^{o}C for 15 mins. A negative-reverse transcriptase control (-RT) was also made using a selected sample, in which the reverse transcriptase was omitted. This was used in reverse transcriptase quantitative polymerase chain reactions (RT-qPCRs) to determine potential DNA contamination.

2.9.3 RT-qPCR procedure

The majority of primer sequences used in RT-qPCRs (see Table 2.4) were from published sources. For occasions when primers were designed, the 'Primer3' program (http://frodo.wi.mit.edu/primer3/) was used to produce primers with an annealing temperature of 60^{o}C and an amplicon size of ≤200bp.

Samples and reagents were loaded into 96- or 384-well qPCR plates (Axygen) with automatic pipettes to minimise error. The cDNA was diluted 1:50 in RNAse-free water, from which 2.5µL was included in a reaction with 5µL SYBR™ Select Master Mix (Thermo-Fisher) and 1.25µM of each primer (Sigma-Aldrich; see Table 2.4). RNAse-free water was added up to a final reaction volume of 10µL. Quantitative gene expression analyses were conducted with the LightCycler® 480 System (Roche) using the following program: 95°C for 2 mins, 50 amplification cycles at 95°C for 15 secs followed by 60°C at 50 secs then a further 10 secs, before a melting curve at 95°C for 5 secs and 65°C for 1 min. Each replicate was loaded twice and only samples for which the duplicates were within 1 'Ct' value were included in

subsequent analyses. A standard curve of each primer pair was calculated, using a serially diluted sample in which the gene was known to be expressed. A primer efficiency from this standard curve of between 1.8-2.2, determined using LightCycler® 480 System Software 1.5 (Roche), was required for consideration. The expression of *Ubiquitin C* (*UbqC*), a ubiquitous house-keeping gene, was determined in all samples and used to normalise the expression of a gene of interest. The mean and the standard deviation were then calculated from normalised triplicate samples. Statistical significance was determined using an unpaired and two-tailed t-test. Two controls for contamination, the –RT from cDNA synthesis and a non-template control, wherein cDNA was omitted from the well in the qPCR plate, were also included. If these returned equivalent 'Ct' values as experimental samples then any source of contamination was rectified and the analyses repeated.

Table 2.4 – Sequences of primers used for RT-qPCR

Gene	Forward sequence	Reverse sequence	Annealing Temp	Reference
Apobec3	CTGGAGCCCCTGTTTCGAAT	CGTATAGGTCCATGGCAGCC	60°C	*Self-designed*
Auts2	GCCGAAGACATCCAACCCTA	AGCTTGTGTGGGTCTGACTG	60°C	*Self-designed*
Dazl	TCTTTGCCAGATATGGCTCAGT	CTTCTGCACATCCACGTCATTA	60°C	Designed by Ruchi Shukla (Meehan group)
Dnmt1	CTGGAAGAGGTAACAGCGGG	CTGGTGTGACGTCGAAGACT	60°C	Designed by Ruchi Shukla (Meehan group)
Dnmt3a	CCTGCAATGACCTCTCCATT	CAGGAGGCGGTAGAACTCAA	60°C	(Ficz *et al.*, 2013)
Dnmt3b	TGGTGATTGGTGGAAGCC	AATGGACGGTTGTCGCC	60°C	(Ficz *et al.*, 2013)
Dnmt3l	ATGGACAATCTGCTGCTGACTG	CGCATAGCATTCTGGTAGTCTCTG	60°C	(Ficz *et al.*, 2013)
Dppa3/ Stella	GCTAACCCTAAACCCCGGTGT	CAATGCGGTTCCGTAGACTGC	60°C	(Huang *et al.*, 2017)
Pdk1	GGCGGCTTTGTGATTTGTAT	ACCTGAATCGGGGGATAAAC	60°C	*Self-designed*
UbqC	GAGTTCCGTCTGCTGTGTGA	TCACAAAGATCTGCATCGTCA	60°C	(Christophorou *et al.*, 2014)
Uhrf1	GCTCCAGTGCCGTTAAGACC	CACGAGCACGGACATTCTTG	60°C	(Ficz *et al.*, 2013)
Utf1	GACTCTGCCTACTTACCGCC	CATGAGTTGGCGGATCTGGT	60°C	*Self-designed*

2.10 Quantification of 5-methylcytosine

Liquid chromatography-mass spectrometry (LC-MS) was employed for the quantification of global DNA methylation levels. This is an unbiased method to assess DNA methylation however, it is unable to detect the genomic redistribution of the mark. Determining the possible biological impact on small albeit statistically significant changes in levels also requires an alternative approach that provides data at the resolution of individual loci.

2.10.1 DNA extraction

The Qiagen DNeasy Blood & Tissue Kit was used in accordance with the manufacturer's instructions to collect DNA for global DNA methylation analysis. All recommended steps were followed, including incubation with RNaseA (Qiagen).

2.10.2 DNA digestion

An aliquot of 2.5µg of DNA was denatured by heating to 95^{o}C for 10 mins in a total volume of 44µL with nuclease free H_2O. Next, 10 U of T7 DNA polymerase (Thermo-Fisher) and 1X reaction buffer was added and incubated overnight at 37^{o}C. The activity of the polymerase promotes the hydrolysis of the DNA strands into single nucleotides. The following day, samples were centrifuged for 45 mins at 12,000 x *g* and 20µL of the supernatant was recovered and added to a new tube, into which 50µL MeOH and 20µL acetonitrile was added. After 10 mins of centrifugation at 17,000 x *g*, the supernatant was recovered, transferred to a new tube and the DNA precipitated in a DNA SpeedVac™. The pellet was then suspended in 20µL of mass spectrometry grade H_2O. Samples were loaded and mass spectra acquired by Jimi Wills (IGMM, University of Edinburgh).

2.10.3 Data analysis

Data was analysed using Xcalibur™ software (Thermo-Fisher) with a processing method designed by Andy Finch (IGMM, University of Edinburgh). The intensity of 5mC as a proportion of the total intensities of cytosine, 5mC and 5hmC from three technical replicates was calculated using this processing method. Error bars were calculated as the standard deviation from the mean. Statistical significance was determined using unpaired two-tailed t-tests. A summary of the molecular masses of the nucleotides searched for in the spectra, as designated on the metabolite database Metlin (https://metlin.scripps.edu/), are presented in Table 2.5. Each individual experiment was performed with internal controls: J1

and TKO-212 (Tsumura *et al.*, 2006) mESCs. Retention times of the nucleotides could then be adjusted accordingly and applied to all samples in the same experiment.

Table 2.5 – Masses inputted to quantify nucleotide abundance

Nucleotide	Mass (m/z)
5-hydroxymethylcytosine	336.0602
5-methylcytosine	320.0653
Cytosine	306.0497

2.10.4 Optimisation of method

In initial attempts to quantify global 5mC levels, DNA was prepared using DNA degradase (Zymo Research) and a 150 x 2.1mm SeQuant® ZIC®-pHILIC column (Millipore) for chromatography. However, the chromatograph peaks of the nucleotides were less defined as they had tails. This introduced subjectivity into the analysis when selecting the area of the chromatograph peak to quantify. Use of the T7 polymerase and a 30 x 1mm HyperCarb column (VWR) column reduced this as demonstrated in Figure 2.1, which shows a representative trace of a J1 mESC sample. Also note the shorter retention time.

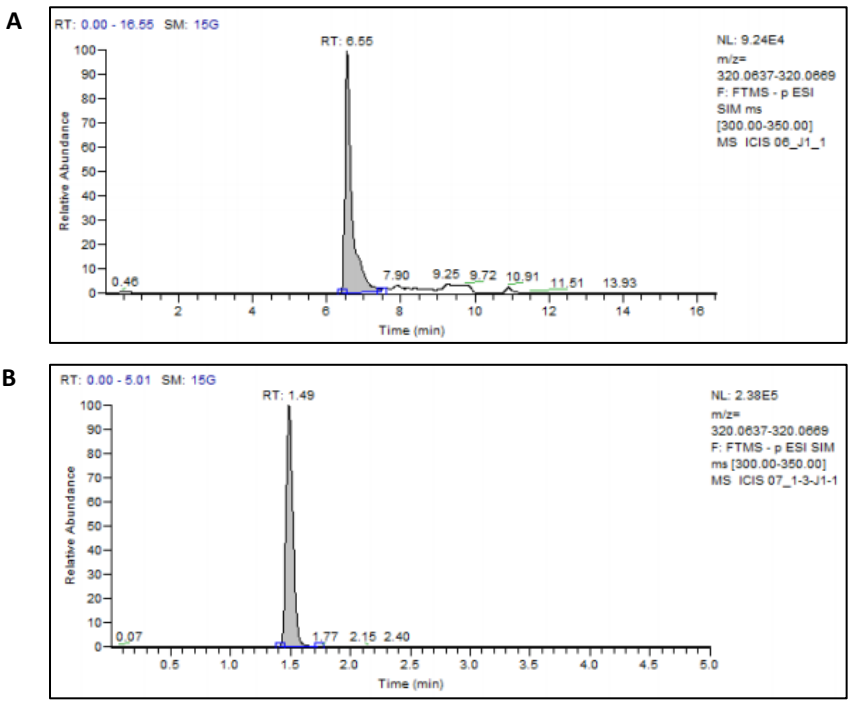


Figure 2.1 - *Figure legend on next page*

Figure on previous page **Figure 2.1** – Representative chromatographs of 5mC using different digestion methods and chromatography columns. **A**. DNA prepared using DNA degradase and chromatography with 150 x 2.1mm SeQuant® ZIC®-pHILIC column. **B**. DNA prepared with T7 polymerase and chromatography with HyperCarb column. Note the less defined peak in **A** compared to **B**.

2.11 Immunofluorescence procedure

Cells were seeded at densities of between 3.0 x 10^5 - 6.0 x 10^5 cells in 12 - or 6-well plates on EtOH sterilised coverslips coated with 0.1% (w/v) gelatin. When appropriate, media was removed and cells were washed 3 X with PBS containing calcium and magnesium (Life Technologies) then fixed for 10 mins at room temperature with 4% (w/v) PFA (paraformaldehyde; Sigma). Blocking was conducted overnight at 4^{o}C in 10% (v/v) donkey serum (Sigma) with 0.1% (v/v) Triton X-100 (Sigma) in PBS. Primary antibodies were diluted appropriately (see Table 2.3) in 1% donkey serum with 0.1% Triton X-100 in PBS and added overnight at 4^{o}C to the coverslips. After 3 X washes with PBS, incubating for 10 mins each time, secondary antibodies (see Table 2.3) in 1% donkey serum with 0.1% Triton X-100 in PBS were applied for 1 hour at room temperature. Coverslips were washed 3 X with PBS and mounted onto glass slides with VECTASHIELD® Antifade Mounting Medium with DAPI (4',6-diamidino-2-phenylindole; Vector Laboratories). Images were acquired on an upright epifluorescence microscope (Zeiss), ensuring exposure times were constant for each channel across samples.

Chapter 3 – Design of a screen to investigate mechanisms of DNA methylation reprogramming in pluripotent cells

3.1 Introduction

DNA hypomethylation is promoted in 2i media however, the mechanism through which this occurs was previously unknown. In this chapter, I have characterised alternative 2i systems and described the set-up of a proteomic screen that aimed to identify candidate regulators of the DNA methylation maintenance machinery.

The emergence of 2i media was the result of attempts to improve traditional serum-based culture media to an extent that mESCs could be maintained *in vitro* with a transcriptional state that more closely recapitulates their *in vivo* counterparts. Further studies noted that 2i promotes DNA methylation reprogramming in mESCs and the distribution resembles patterns in cells of the ICM of early embryos. As a consequence, 2i provides an opportunity to understand the role of DNA methylation within early development and decipher the molecular mechanisms that regulate this epigenetic process. The effects of 2i are also influenced by serum withdrawal or depletion which has led to various iterations of the base medium being conceived (Ying *et al.*, 2008; Ficz *et al.*, 2013; Habibi *et al.*, 2013; Leitch *et al.*, 2013; Hackett *et al.*, 2017).

The vast majority of publications utilise N2B27 media, a system used previously to sustain neural populations in culture (Ying *et al.*, 2008).There have also been reports that KOSR can directly substitute for the FCS that supplements GMEM in standard mESC culture and is a sufficient base media for 2i to exert an effect. In these studies, this composition, referred to hereafter as KOSR+2i, can drive the establishment of iPSCs from a partially reprogrammed starting state. This is characterised, amongst other markers, by the expression of endogenous *Oct4* at the end of the reprogramming process, rather than a dependency on exogenous *Oct4* initially transduced with a viral vector (Silva *et al.*, 2008; Christophorou *et al.*, 2014). The KOSR reagent has also been reported to sustain human ESCs (Garcia-Gonzalo and Izpisúa Belmonte, 2008) in addition to primordial and embryonic germ cells (Horii *et al.*, 2003).

Transcriptomic analyses conducted on mESCs treated with 2i in either a KOSR or N2B27 base medium unsurprisingly highlight differences between the systems. Importantly however, the profiles of cells in the serum depleted medias have more similar gene expression profiles than a population cultured in traditional media. Globally, transcripts are more reflective of an earlier naïve state in development, contrasting with the serum culture which is heterogenous and markers associated with differentiation are also expressed (Hackett *et al.*,

2017). This demonstrates that the 2 molecular inhibitors are largely able to overcome the effects of other ingredients in the base medias.

In this results chapter, I detail the set-up of a screen to identify mechanisms which may contribute to the reprogramming of DNA methylation. By transitioning mESCs from serum-based culture to both of the aforementioned 2i systems in parallel, I characterised the kinetics of DNA demethylation. The 2i system with the greatest dynamic range, with respect to DNA methylation levels relative to serum-based culture, will then be taken forward and used to screen for interacting candidates of DNMT1 and Uhrf1. Using the media that imposes the greatest change will facilitate the quantification of any effects on DNA methylation in functional studies of the candidates.

3.2 Results

3.2.1 Characterising DNA methylation in KOSR and N2B27 2i media systems

To establish the effects of the two different culture conditions on DNA methylation, I performed a time course experiment in which global methylation levels were assessed, in addition to the gene and protein expression of DNA methylation regulators. To this end, I conducted LC-MS for the quantification of global 5mC, RT-PCR for gene expression analyses and western blotting to assess protein levels. Samples were taken daily for five days and another at day 14 in culture. This time point has previously been reported as the time when cells reach a steady state with respect to methylation levels (von Meyenn *et al.*, 2016).

For each LC-MS experiment, two control cell lines were also included: TKO mESCs, in which all three *Dnmts* have been knocked out, and the parental J1 line (Tsumura *et al.*, 2006). These provided an internal control when quantifying unknown levels of DNA methylation. The values for each of the daily time points are presented relative to that obtained from WT E14 cells cultured in serum, which is taken as 1. As the experiment commenced with cells from the same line, the starting methylation state for both KOSR+2i and N2B27+2i conditions is equivalent.

Within 1 day of transitioning cells from serum culture, both of the 2i medias begin to promote hypomethylation. The effect however is more pronounced in KOSR+2i throughout the time course. The lowest level of methylation attained in N2B27+2i culture, a 30% loss after 14 days relative to serum at day 0, is reached within 2 days of KOSR+2i application. By contrast, after 14 days in KOSR+2i, cells exhibited a loss of methylation of over 80%. This demonstrates that more efficient demethylation is occurring in the latter system (Figure 3.1A). In order to quantity this, the relative values from days 0-5 in culture were transformed into their natural logs and plotted against time to enable the rate of demethylation to be calculated. Assuming first order kinetics, the slopes of the respective trendlines show that KOSR+2i promotes the loss of DNA methylation at an approximately 4-fold faster rate, compared to N2B27+2i, a difference of high statistical significance ($p \leq 0.001$) (Figure 3.1B).

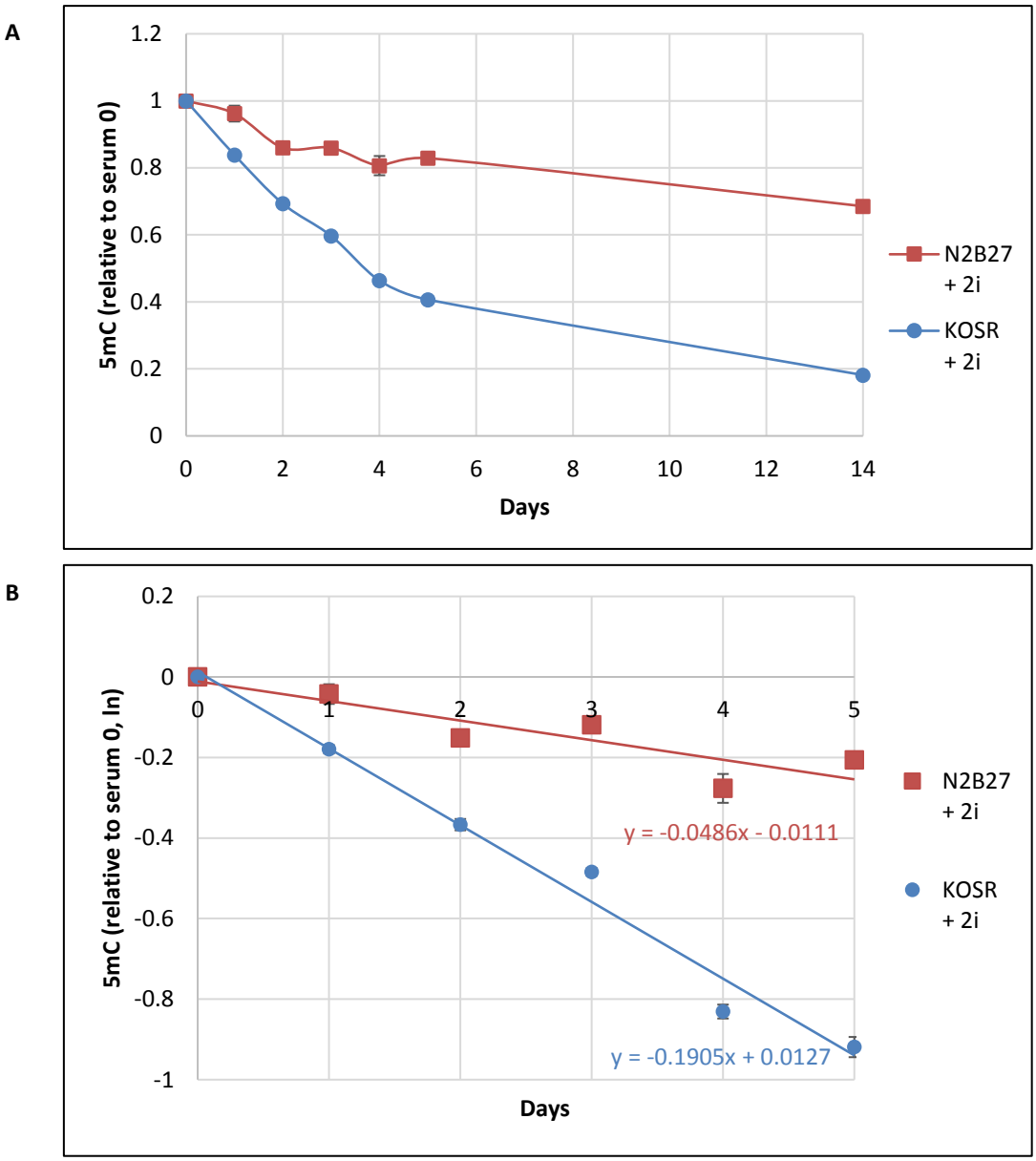


Figure 3.1 – A. DNA methylation levels of WT E14 mESCs cultured in different media, KOSR and N2B27, with 2i for the days indicated. Values presented as relative levels to serum cultured cells at 0 days. **B**. The relative values from days 0-5 of culture presented as natural logs. The slope of the line indicates the rate of demethylation, assuming first order rate kinetics. Error bars represent the standard deviation of the mean from three technical replicates. Representative experiment from two repeats.

3.2.2 Protein and mRNA dynamics of DNA methylation mediators in KOSR and N2B27 2i systems

To assess the impact of the different 2i media compositions on the protein levels and expression of the DNA methylation machinery, I performed western blots and RT-qPCRs as described above. In contrast to the serum and N2B27 + 2i culture systems, Dnmt1 protein was subject to a partial downregulation in KOSR + 2i after 3 days. Contrary to previous publications (von Meyenn *et al.*, 2016), Uhrf1 protein levels were shown to not be affected in N2B27 + 2i, similarly to serum, whereas a decline was precipitated after 2 days in KOSR + 2i. It is important to note that a relative drop of methylation by 30% was recorded in KOSR+2i medium prior to the loss of Uhrf1. In both 2i systems, Dnmt3a was regulated similarly as levels dropped sharply by 3 days. Meanwhile Dnmt3b protein was present in KOSR+2i, albeit to a lower extent than serum. Its regulation in N2B27+2i mirrored Dnmt3a whereby a steep decline was seen (Figure 3.2).

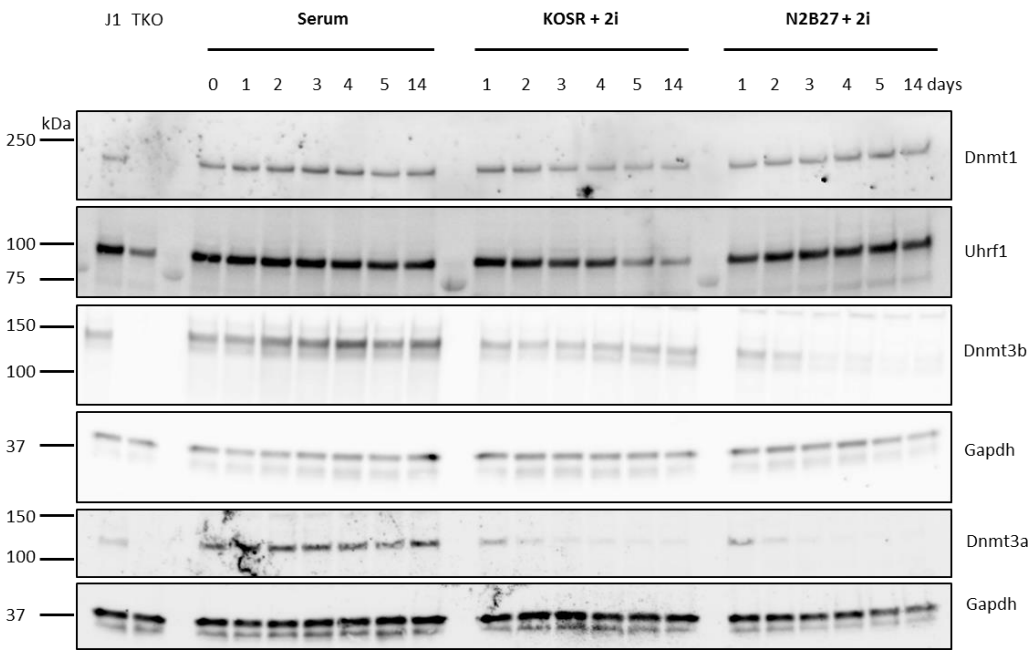


Figure 3.2 – Western blot of whole cell lysates from cells cultured in traditional media supplemented with FCS (serum) or two types of 2i (KOSR + 2i, N2B27 + 2i). Time since switching to 2i culture represented as days. Gapdh was used as a loading control and TKO and parental J1 lysates are shown to display the specificity of the antibodies. Representative of two independent experiments.

Comparison of *Dnmt1* gene expression between the different 2i media shows it is subject to fluctuation throughout culture however, it reaches a state comparable to that of serum-maintained cells in the latter stages of the transition. Despite the mRNA of *Uhrf1* being 2-fold higher in KOSR+2i by day 14, the protein was still evidently subject to posttranslational regulation. At the end of this time course, the N2B27+2i media also displayed higher levels of *Uhrf1* mRNA compared to serum cultured cells yet protein was sustained to a comparable level (Figures 3.2 and 3.3).

The downregulation of Dnmt3b protein in N2B27+2i can be accounted for by the rapid drop in transcript levels. While a decline was initiated in KOSR+2i, the mRNA began to rebound after 5 days and remained comparable to the serum culture thereafter, despite protein levels being lower. This direct relationship to mRNA and protein also holds for Dnmt3a which in both treatments declined rapidly. Statistically significant ($p \leq 0.05$) differences are evident at some time points, particularly in the earlier stages of the transition. In all occasions however, levels remain lower than in serum (Figures 3.2 and 3.3).

Interestingly, whilst *Dnmt3l* is almost undetectable in N2B27+2i culture after 2 days, expression increases after 3 days in KOSR+2i despite an initial drop, which mirrors the regulation of *Dnmt3b* in the same system (Figure 3.3).

Of note, despite the intention of only including it as control to validate the Dnmt antibodies, the TKO line is shown to also have a lower amount of Uhrf1 protein, suggesting an association between Dnmts and Uhrf1 stability (Figure 3.2).

Based on the results above, and as discussed later, KOSR+2i represented the most informative system to use for investigating mechanisms that lead to hypomethylation. This is due to the more pronounced demethylation following 14 days of culture and the early steep decline compared to N2B27+2i.

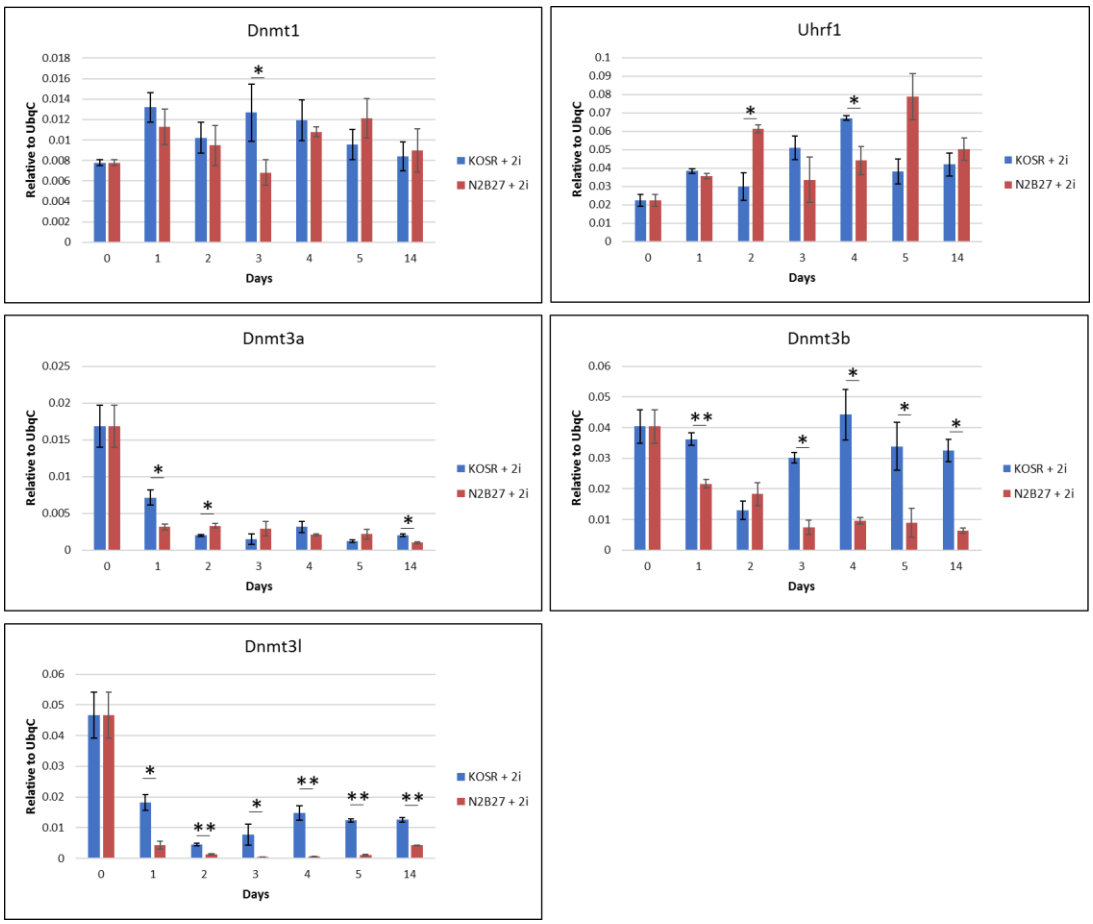


Figure 3.3 – RT-qPCR analysis of DNA methylation mediators in two different 2i media compositions, KOSR + 2i and N2B27 + 2i, at the indicated time in days. Values are presented relative to *UbqC* expression. Error bars represent the standard deviation from three technical replicates. (* = $p \leq 0.05$; **= $0.001 < p < 0.05$). Representative of two independent experiments.

3.2.3 Generation of cell lines expressing DNMT1 and Uhrf1 -BirA fusion proteins*

To elucidate routes by which hypomethylation is promoted, I sought to characterise the interactome of the fundamental components of DNA methylation maintenance, DNMT1 and Uhrf1. Capturing the interacting partners of these proteins in mESCs maintained in serum-based media and KOSR+2i, where DNA methylation levels are markedly different, could lead to the identification of novel regulators. For example, if an enzyme interacts more prevalently with Uhrf1 in KOSR+2i compared to serum, it could be inferred that Uhrf1 is impaired by a PTM. Likewise, if a complex forms with Dnmt1 in serum media, this could be hypothesised to promote Dnmt1 activity. This dataset would ultimately facilitate the design of functional studies to determine whether the interacting partners influence DNA methylation

maintenance. The method of choice for identifying the interacting partners of DNMT1 and Uhrf1 was BioID, owing to the advantages over Co-IPs as discussed below.

The BioID method utilises a promiscuous biotin ligase, BirA*, that modifies lysine residues of vicinal proteins that are within a 10 nm radius. If BirA* is tagged to a protein of interest, biotinylated proteins can therefore be inferred as interacting partners or within a macromolecular complex with the protein of interest. It is necessary for cell culture media to be supplemented with biotin for a period of up to 24 hours. As biotinylation is a rare biological process, there is a high degree of confidence that biotinylated proteins are within close proximity or indeed interactors of the tagged protein in cells. To recover the biotinylated proteins, harsh lysis conditions can be used as there is no requirement to preserve the interaction. This is particularly useful for DNMT1 and Uhrf1 which are chromatin associated factors. Furthermore, as the pull-down is mediated by streptavidin, stringent conditions can be used to remove background proteins as the interaction between biotin and streptavidin is one of the strongest known in nature. Proteins that are pulled-down in a BioID experiment can then be identified by mass-spectrometry. Those that are enriched by a particular factor over control conditions as outlined below are then taken forward for consideration (Figure 3.4) (Roux *et al.*, 2012; Varnaitė and MacNeill, 2016).

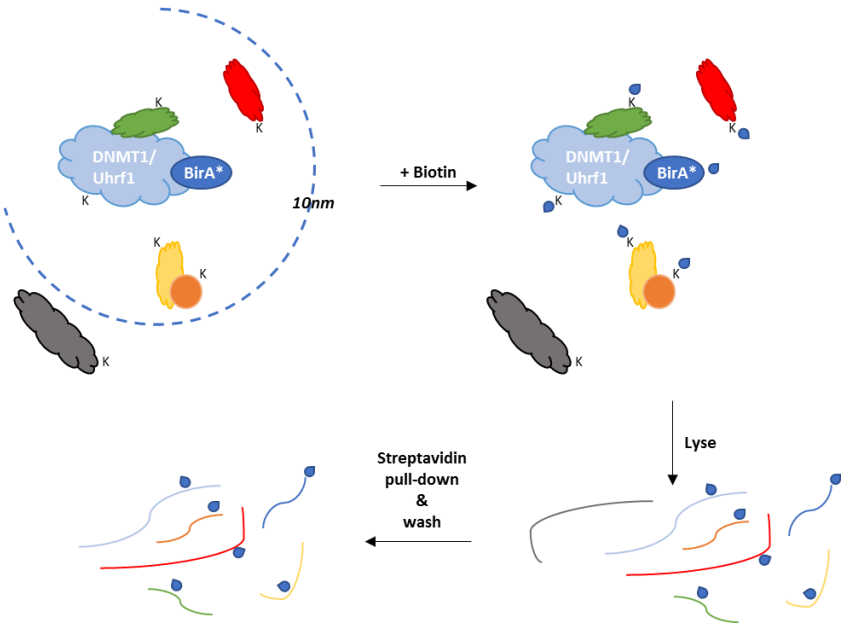


Figure 3.4 – *Figure legend on next page*

Figure on previous page **Figure 3.4** – Overview of the BioID method. A promiscuous biotin ligase (BirA*) is fused to proteins of interest, such as DNMT1 and Uhrf1. If vicinal proteins are within 10nm then they are susceptible to biotinylation (blue teardrops) at lysine residues (K) when biotin is provided exogenously. Cells are then lysed and the biotin moiety is maintained, enabling streptavidin mediated pull-down. Following this, the recovered proteins can be identified by mass-spectrometry. Adapted from (Roux *et al.*, 2012).

Importantly, equivalent cell lysis and washing steps would require extensive optimisation in order to capture the interactome with a conventional co-IP. Lysis conditions to solubilise chromatin associated factors and the subsequent washing steps of an appropriate stringency need to be balanced against the potential disruption of interactions between the target of interest and other proteins. In this regard, the use of BioID is advantageous (Roux *et al.*, 2012; Varnaitė and MacNeill, 2016).

To establish the requisite experimental cell lines for BioID, I cloned human *DNMT1* and mouse *Uhrf1* into BirA* expressing vectors and stably transfected these into an E14 cell line (Figure 3.5). A more complete overview of this process is described earlier in Chapter 2 which covers both the initial cloning of *BirA** into a piggyBac vector and the subsequent In-Fusion® mediated cloning of the genes of interest. I will reiterate here however that as both DNMT1 and Uhrf1 are large multidomain proteins, lines were generated so that the proteins were tagged at the N- and C-terminus in order to capture a broader number of partners that may specifically interact with one portion of the protein.

To validate the cell lines, several assays were performed. Firstly, I confirmed that the tagged protein is expressed by western blotting. As expected, both an endogenous band was present and another ~30kDa higher, corresponding to the protein fused with BirA*, for both the DNMT1 and Uhrf1 lines. Whilst Uhrf1-BirA* is expressed at a level lower than the endogenous protein, the DNMT1 fusion protein is more abundant than endogenous Dnmt1. Two control lines were also generated: an empty vector control, which expresses the piggyBac vector, and a free BirA* control that is not fused or tethered to any protein (Figure 3.6). The BirA* line represents the most stringent control as this accounts for proteins that may interact with the fusion proteins by virtue of the additional tag. The Empty vector control accounts for proteins that bind to the streptavidin-conjugated beads irrespective of whether they are subject to biotinylation.

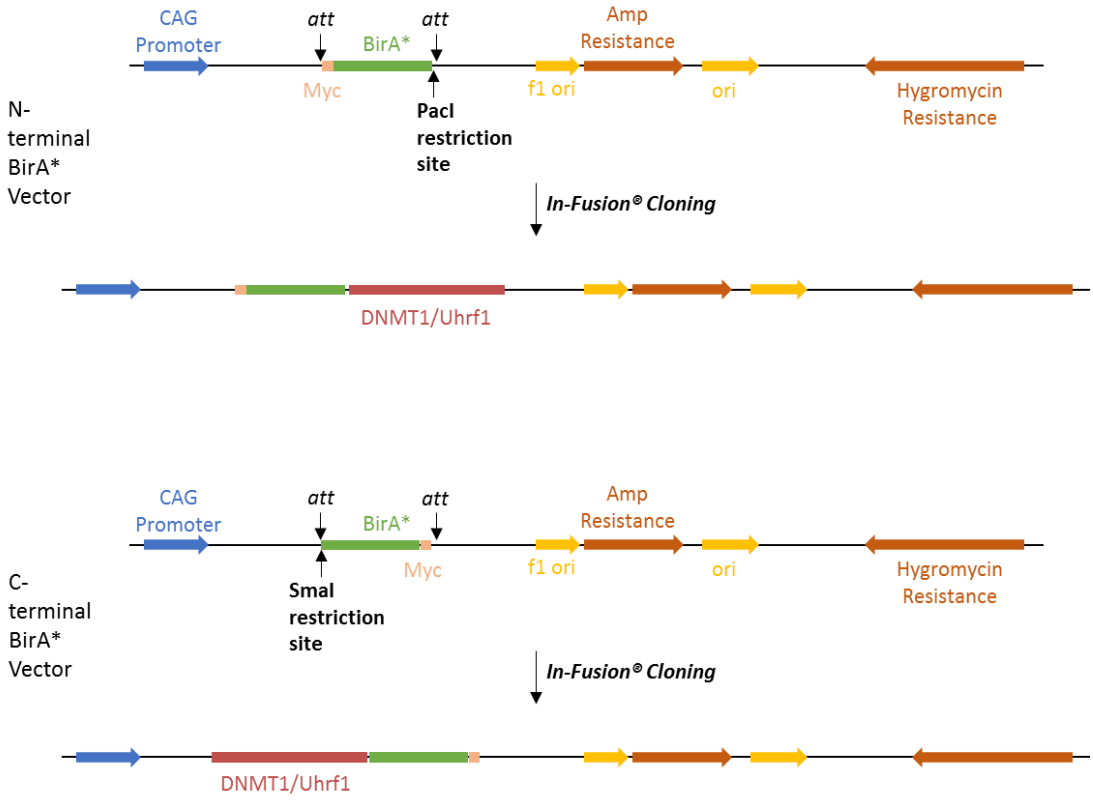


Figure 3.5 – Linear maps of *DNMT1-* and *Uhrf1-BirA** vectors. Expression is regulated by a CAG promoter. Ampicillin and Hygromycin resistance genes permit selection of positive bacterial clones and successfully transfected cells respectively

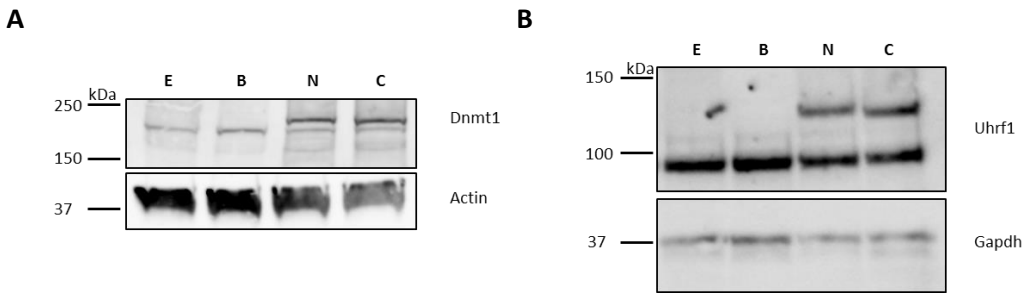


Figure 3.6 – Western blot of whole cell lysates expressing **A** Empty control (E), Free BirA* control (B), BirA* N-terminally tagged DNMT1 (N), BirA* C-terminally tagged DNMT1 (C) vectors and **B** BirA* N-terminally tagged Uhrf1 (N) and BirA* C-terminally tagged Uhrf1 (C) vectors. Upper bands in **A** Dnmt1 and **B** Uhrf1 blots represent BirA* fusion proteins and the lower band the endogenous proteins. Actin and Gapdh were used respectively as loading controls.

As depicted in the vector maps (Figure 3.5), the BirA* protein possesses a Myc-tag which enables the expression levels of the tagged proteins and the free BirA* to be evaluated. A western blot was performed following a time-course wherein biotin was applied to the lines for 0, 2, 4, 6 and 24 hours. As shown in Figure 3.7, the expression of the free BirA* is comparable to that of the DNMT1-BirA* fusions. The N-terminally tagged protein however appears to be subject to cleavage. Figure 3.8 demonstrates that free BirA* is more abundant than both Uhrf1-BirA* proteins. Ideally, the levels would be similiar as this would prevent candidates that are true interacting partners from being discarded ; free BirA* may interact more frequently with other proteins if there is more of it. Therefore with respect to the Uhrf1-BirA* lines, candidates that are indeed enriched over the free BirA* control can be considered interacting proteins with a high degree of confidence. Similarly, due to the deficiencies of the the N-termially tagged DNMT1 line, candidates enriched over the free BirA* control would also be strong contenders. There initially appears to be a decrease in DNMT1- and Uhrf1-BirA* proteins at 24 hours however, it should be noted that the Myc-tag has a lysine residue and this could be subject to biotinylation. As a consequence the epitope of the Myc antibody may be partially obscured.

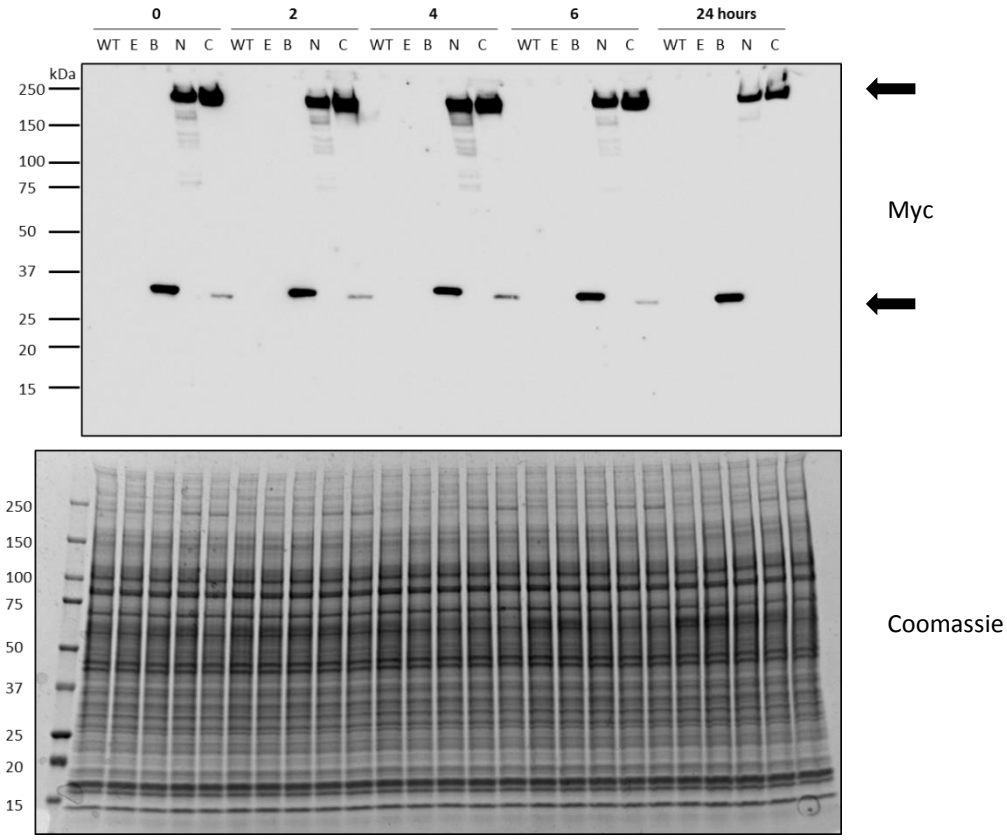


Figure 3.7 – Western blot of whole cell lysates to assess the expression levels of N and C-terminally tagged DNMT1-BirA* proteins (upper arrow) and free BirA* (lower arrow). The presence of free BirA* in the DNMT1-BirA* lanes is indicative of sample degradation or cleavage of the BirA* tag. Coomassie staining of a gel was used as a loading control. Empty vector control (E), free BirA* control (B), BirA* N-terminally tagged DNMT1 (N) , BirA* C-terminally tagged DNMT1 (C)

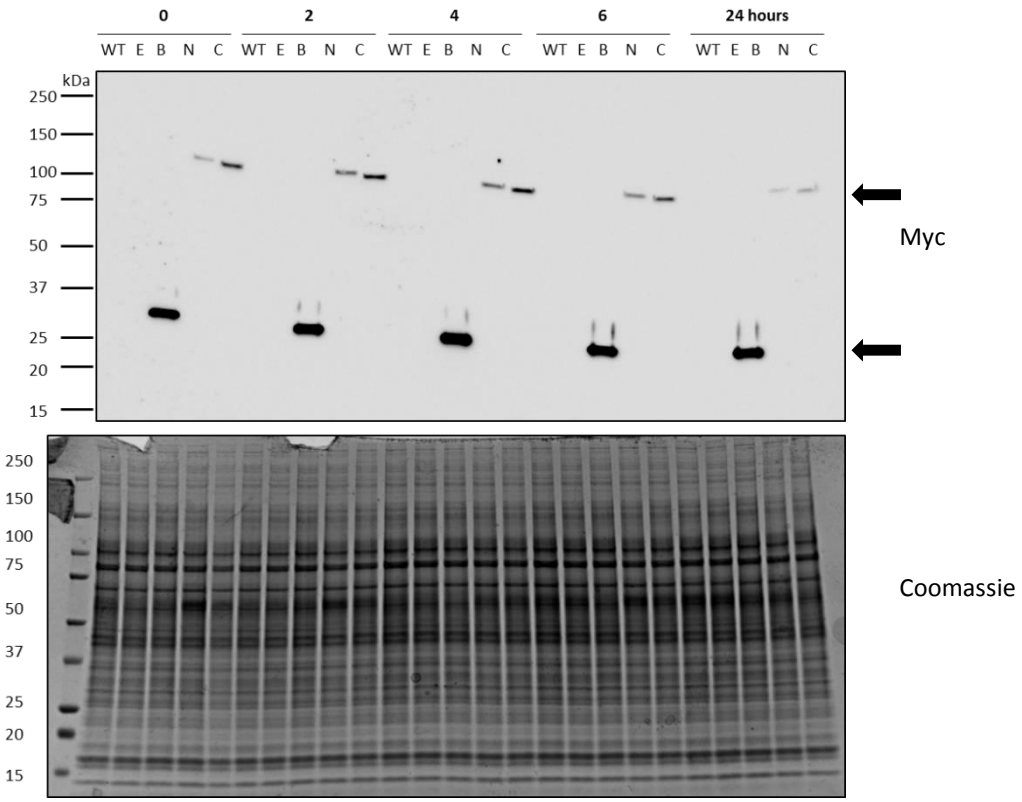


Figure 3.8 – Western blot of whole cell lysates to assess the expression levels of N and C-terminally Uhrf1-BirA* proteins (upper arrow) and free BirA* (lower arrow). Coomassie staining of a gel was used as a loading control. Empty vector control (E), free BirA* control (B), BirA* N-terminally tagged Uhrf1 (N) and BirA* C-terminally tagged Uhrf1 (C).

3.2.4 Testing catalytic activity of BirA fusion proteins*

Although the fusion proteins are expressed, it was also important to determine that the BirA* is functional. To this end, the lines were treated with biotin for 0, 2, 4, 6 or 24 hours and whole cell extracts prepared and analysed by a western blot. An antibody conjugated to streptavidin was then applied to detect all biotinylated proteins. As shown in Figure 3.9, the DNMT1 lines were indeed shown to be functional, with the extent of biotinylation increasing with time. Also evident is the relatively greater acitivty displayed by the free BirA*. As DNMT1-BirA* and free BirA* are expressed at similar levels (see Figure 3.7), this would suggest that the localisation of the proteins, enabling it to find more substrates, can account for this difference. Another non-mutually exclusive posibility is that the free BirA* is

intrinsicly more active. Figure 3.10 initially appears to suggest that the Uhrf1-BirA* line has either weak biotinylation activity or is non-functional.

Extensive autobiotinylation of the exogenously introduced Uhrf1 and DNMT1 is also expected. Blots for the Myc tagged protein after biotin treatment and streptatividin pull-down can therefore also serve as indicators that the BirA* is functional. As shown in Figures 3.11 and 3.12, both proteins are in the lanes corresponding to the pull-down suggesting autobiotinylation and that the BirA* fusion proteins have biotinylating activity. The low level expression of Uhrf1-BirA* (see Figure 3.8) may therefore account for the inability to detect biotinylated proteins in Figure 3.10. Also shown in Figures 3.11 and 3.12 is an enrichment of biotinylated proteins in the lanes corresponding to the pull-down. In order to validate the suitability of the lysis conditions, the cell debris that remains after clarifying the lysate prior to performing the pull-down has also been included in these blots (Pellet). For both DNMT1- and Uhrf1-BirA*, very little protein is detected in the Pellet. Therefore, it is evident that the vast majority has been solubulised, as have potential chromatin associated interacting partners.

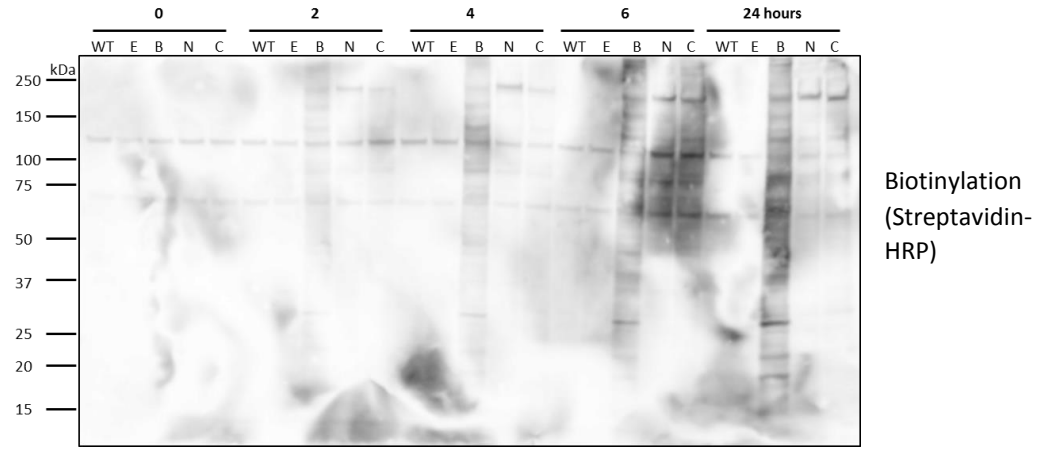


Figure 3.9 – Western blot of whole cell lysates to determine the biotinylation function of N and C-terminally tagged DNMT1-BirA* proteins following application of biotin for the time indicated. Wild-type E14 (WT), Empty vector control (E), free BirA* control (B), BirA* N-terminally tagged DNMT1 (N) , BirA* C-terminally tagged DNMT1 (C)

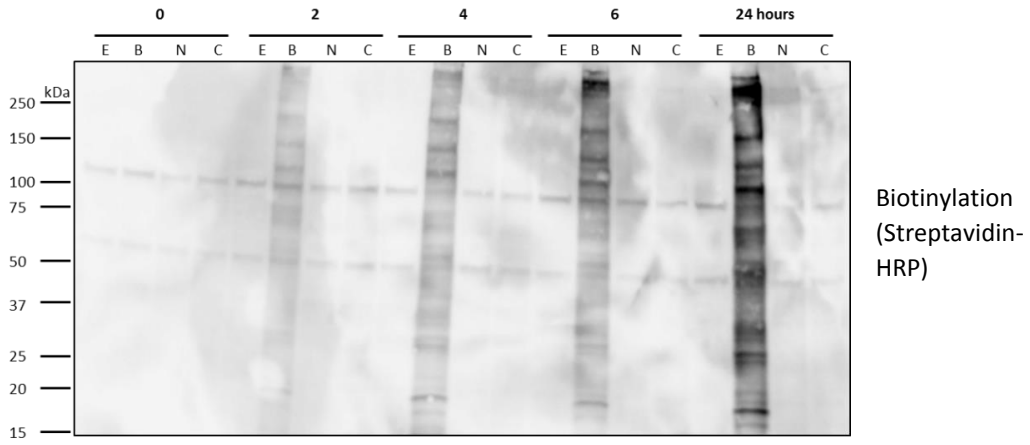


Figure 3.10 – Western blot of whole cell lysates to determine the biotinylation function of N and C-terminally tagged Uhrf1-BirA* proteins following application of biotin for the time indicated. Empty vector control (E), free BirA* control (B), BirA* N-terminally tagged Uhrf1 (N) , BirA* C-terminally tagged Uhrf1 (C)

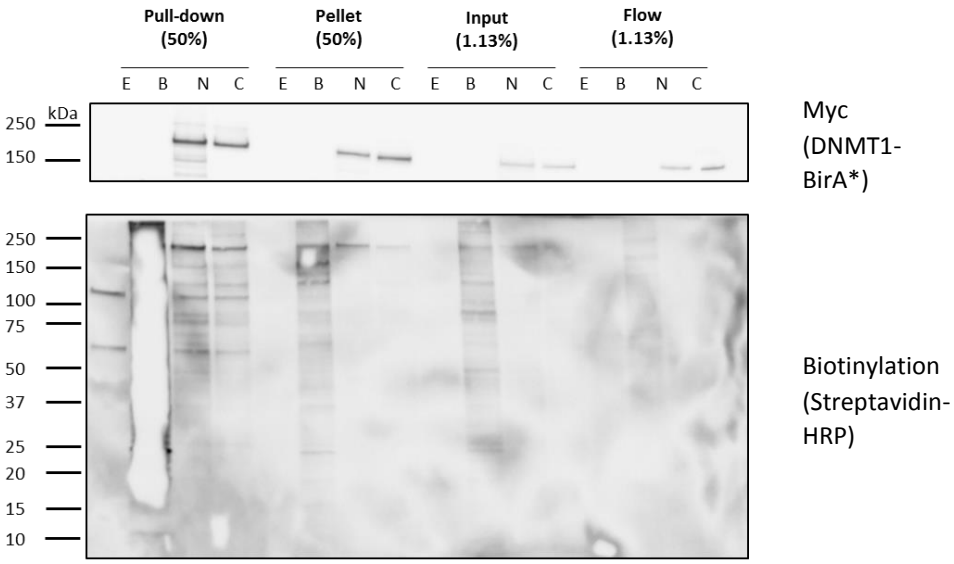


Figure 3.11 – Western blot of whole cell lysates from BirA* fusion protein and control lines following treatment with biotin for 24 hours and streptavidin mediated pull-down. Please note the loading order**.** The pellet heading refers to the cell debris that remains following the clarification of the cell lysate prior to the pull-down**.** Empty vector control (E), free BirA* control (B), BirA* N-terminally tagged DNMT1 (N) , BirA* C-terminally tagged DNMT1 (C).

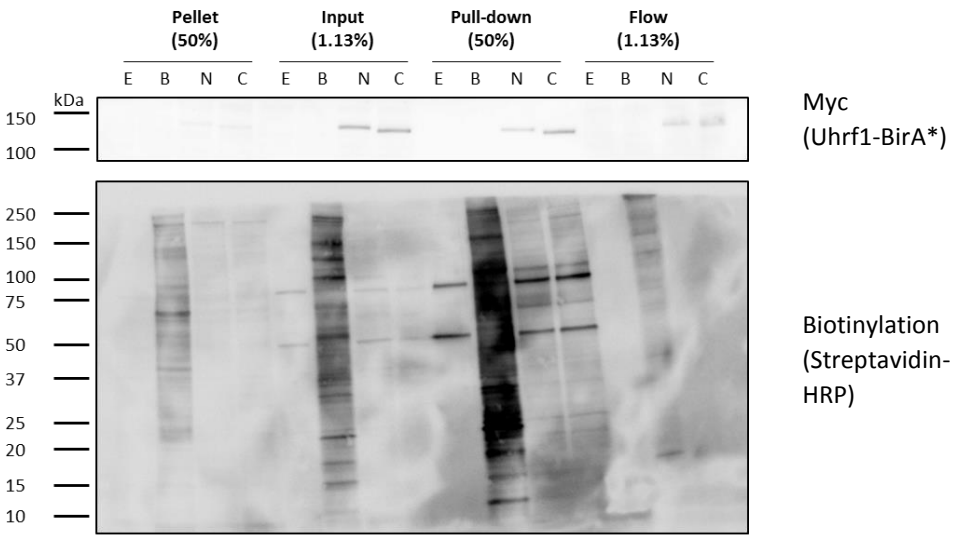


Figure 3.12 – Western blot of whole cell lysates from BirA* fusion protein and control lines following treatment with biotin for 24 hours and streptavidin mediated pull-down. Please note the loading order**.** The pellet heading refers to the cell debris that remains following the clarification of the cell lysate prior to the pull-down**.** Empty vector control (E), free BirA* control (B), BirA* N-terminally tagged Uhrf1 (N) , BirA* C-terminally tagged Uhrf1 (C)

3.2.5 Determining the localisation of BirA fusion proteins*

Another important consideration before embarking on the screen was whether DNMT1- and Uhrf1-BirA* fusion proteins localise to the nucleus, as would be expected of the endogenous proteins. This would increase the confidence that novel interactors found in the screen are not the result of an artefact in the system from mislocalised protein. This was tested by immunofluorescence, exploiting the Myc-tag and by DAPI counterstaining (Figure 3.13). For each line, six separate representative images were taken. Merged channels of DNMT1- and Uhrf1-BirA* with DAPI suggest that the fusion proteins are exclusively localised to the nucleus. Foci of colocalisation (white arrows) are visbile between DAPI dense heterochromatic regions and the Uhrf1-BirA* fusion. However as the DNMT1-BirA* is ubiquitous in the nucleus, similar incidences are not evident in the respective lines. A high density of Uhrf1-BirA* is also detected in at least two cells of a colony in the images taken, as depicted in the N-terminus line and other represenitive examples not shown. This may be associated with cell cycle regulation and will require further investigation using relevant markers. The staining of the free BirA* line reveals that it is prevalent throughout the nucleus

and the cytoplasm, which may account for the extensive biotinylation, compared to the fusion proteins (Figures 3.9 and 3.10).

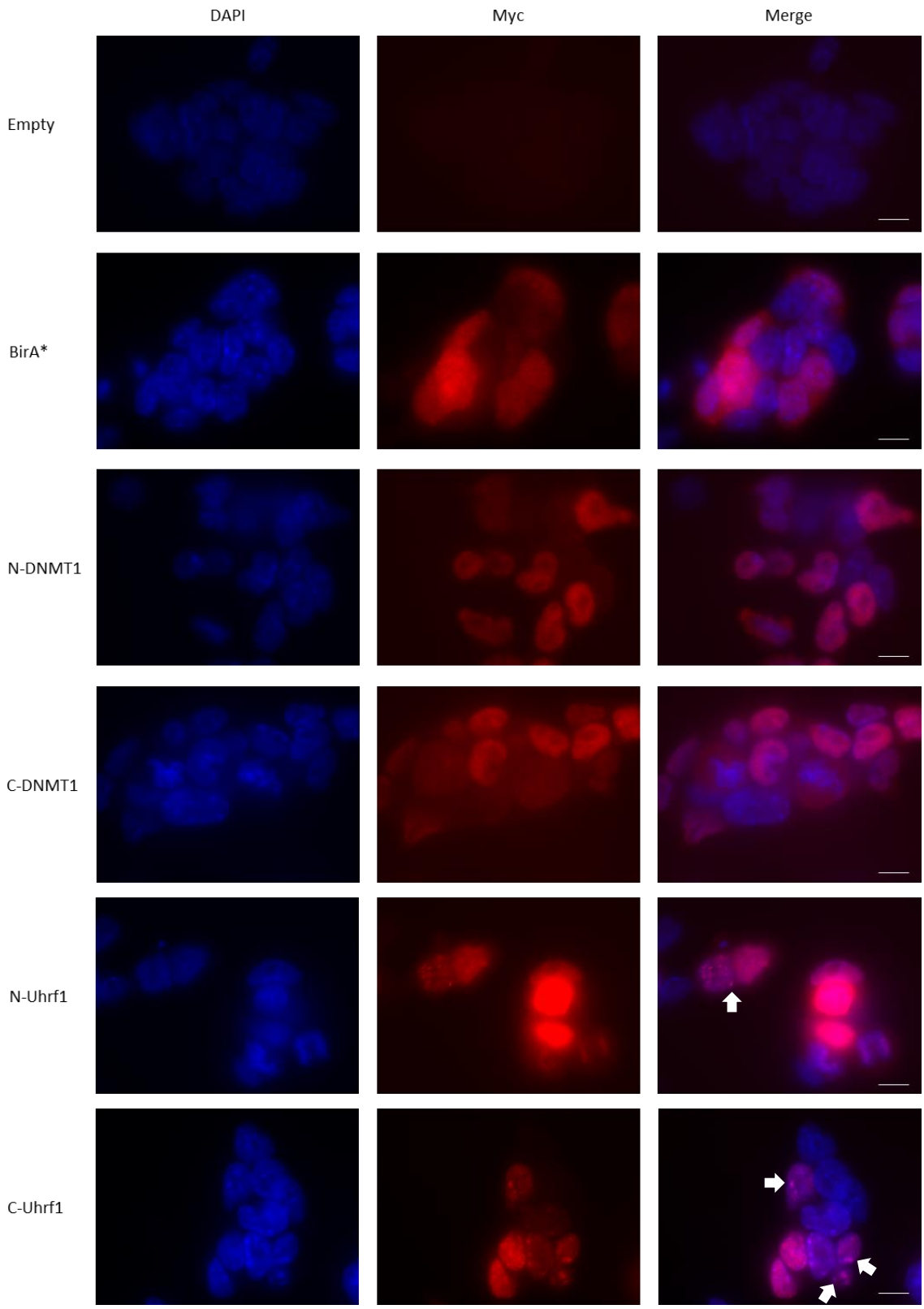


Figure 3.13 – Immunofluorescent staining of BirA* fusion proteins and control lines. Arrows indicate regions of colocalisation with DAPI dense heterochromatic regions. Images representative of six individual colonies. Images acquired with X100 objective. Scale bar = 10 µm.

3.2.6 Developing an independent method to validate the BioID screen

To increase the confidence that a protein identified in the BioID screen is an interacting partner, pull-downs were performed in parallel. Figure 3.14 shows that the Myc-tag can be used to achieve efficient pull-down of the fusion proteins without biotin, in addition to the free BirA* which would serve as the non-specific binding control. This also provides further evidence that the N-terminally tagged DNMT1 protein is cleaved. As a protein interaction is independent of biotin and its supplementation is only required in BioID, it is appropriate to use the N-terminally tagged DNMT1 and Uhrf1 lines only. Although this route may be insufficient to detect transient interactors, proteins identified in both the Myc pull-down and BioID are likely to be robust interactors and would be the strongest candidates to pursue in future functional studies.

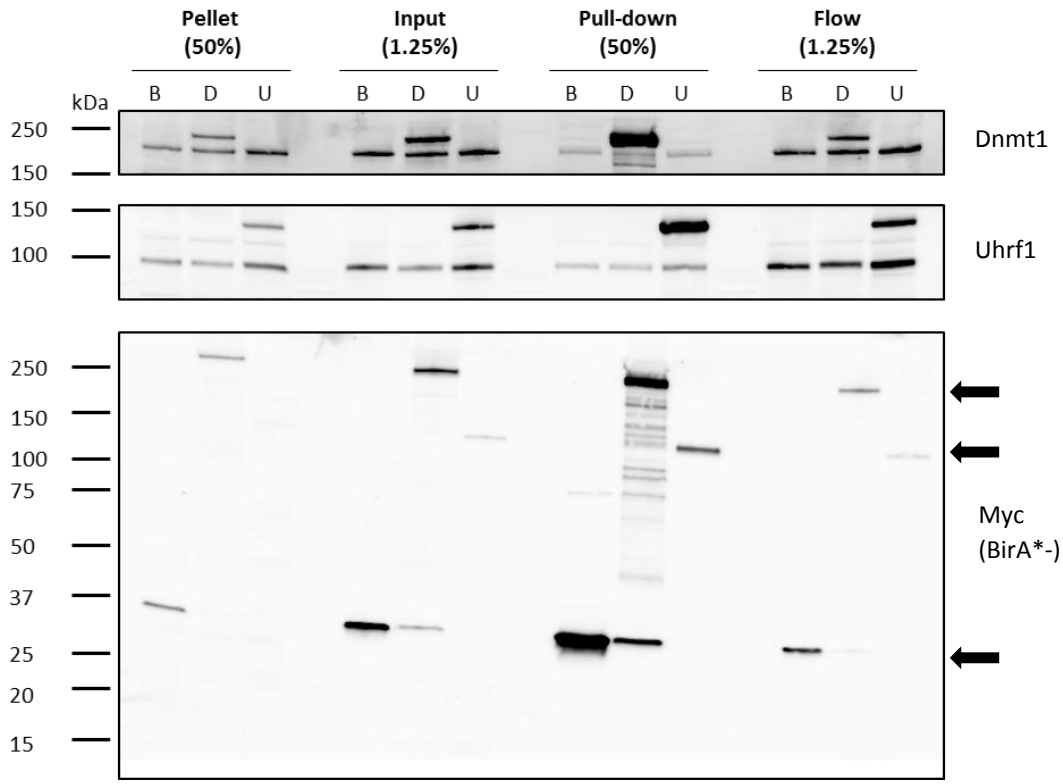


Figure 3.14 – Western blot to show Myc-mediated pull-down of BirA* proteins from the respective stable cell lines. Upper arrow to DNMT1-BirA*, middle to Uhrf1-BirA* and lower to BirA*. The presence of free BirA* in the DNMT1-BirA* lane is indicative of sample degradation or cleavage of the BirA*. Pellet refers to the cell debris remaining after clarifying the cell lysate prior to the pull-down. Free BirA* control (B), BirA* N-terminally tagged DNMT1 (D) , BirA* N-terminally tagged Uhrf1 (U).

3.3 Discussion

Prior to the onset of this study, the major mechanism through which 2i promotes hypomethylation, particularly in the N2B27 based medium, was unknown. However, in 2016, it was reported that downregulation of Uhrf1 precipitates the decline in global methylation, with some additional effects accounted for by the reduction in Dnmt3a and -b levels. This results in impaired maintenance methylation as Dnmt1 is not as pervasively recruited to chromatin and over time with cell division, DNA methylation is diluted. However, marks are maintained at certain regions including major satellites and imprinted genes (von Meyenn *et al.*, 2016). To my knowledge, I have presented the first characterisation of the dynamics of DNA methylation and the mediators in an equivalent system known as KOSR+2i.

3.3.1 KOSR+2i is an appropriate system to investigate mechanisms that regulate DNA methylation

Despite using previously published protocols, I did not observe the extensive hypomethylation that has been reported upon culture of mESCs in N2B27+2i media (von Meyenn *et al.*, 2016). Therefore, the influence of environmental factors, such as the CO_2 pressure, and the use of different cell clones will need to be taken into consideration. I raise the former point as it has been noted that for the duration of my studies, the CO_2 was maintained at lower than 1 bar (this has since been rectified). As the epigenetic reprogramming in 2i based media is mediated by the application of cell signalling modulators, conflicting pathways that are influenced by variable culturing conditions may antagonise the effects that are otherwise observed in a different environment. Nevertheless, rapid hypomethylation was promoted with culture of mESCs in KOSR+2i.

As Uhrf1 was not downregulated in N2B27+2i in the experiments presented above and the effects on DNA methylation were less dramatic, this provides further evidence that the mechanism that accounts for the majority of the hypomethylation in N2B27+2i is indeed mediated through Uhrf1. Although an experiment would need to be conducted in parallel before decisively concluding as such, I hypothesise that the demethylation observed in the N2B27+2i condition above (see Figure 3.1) may largely be accounted for by the downregulation of *Dnmt3a* and *-b*. Reviewing published studies indicates that the rate of demethylation in *de novo Dnmt* null cells is comparable to the rate observed here with N2B27+2i application. Throughout this experiment, the cell lines underwent 6 passages in

both media and it has been reported that *Dnmt3a -/-*; *Dnmt3b -/-* null cells can lose up to 20% of their methylation within 5 passages (Jackson *et al.*, 2004). As the proliferation rate is comparable between serum and N2B27+2i cultured cells (ter Huurne *et al.*, 2017), there may still be a contribution from additional mechanisms acting on the maintenance pathway. As Dnmt1 and Uhrf1 protein levels were consistent, PTMs or a disruption in protein complexes may also be accountable for the 30% drop recorded in N2B27+2i, in addition to *Dmt3a* and -*b* downregulation.

Notwithstanding the effect on Uhrf1, other differences were apparent in the regulation of the *de novo* transferases between the 2i culture systems. Whilst Dnmt3a declined rapidly in both systems, albeit relatively faster at the transcript level in N2B27, Dnmt3b remained expressed to a higher degree at both the mRNA and protein level. One factor previously reported to dictate the expression of the *de novo* transferases in 2i is Prdm14. Prdm14 is an important specification factor of germ cells and its upregulation anticorrelates with Dnmt3a and -b levels (Grabole *et al.*, 2013). Differential regulation of Prdm14 by KOSR and N2B27 based systems may account for this disparity.

With two different base media available, a question that arises is whether an N2B27 or KOSR base is the most appropriate system to investigate the mechanisms of DNA demethylation.

Although it is described as chemically defined, the exact composition of KOSR has not been fully disclosed however, a group has identified numerous components by mass spectrometry (Garcia-Gonzalo and Izpisúa Belmonte, 2008). Like standard FCS, it is also reportedly subject to lot-to-lot variability which may decrease the veracity with which it can model naïve pluripotency (Chaudhry *et al.*, 2008; Tamm, Pijuan Galitó and Annerén, 2013). Of note, is the maintained expression of Dnmt3b in KOSR+2i (Figure 3.2), which has been reported to occur *in vivo* in most stages of pre-implantation development (Hirasawa *et al.*, 2008).

In my current attempts, I have been unable to recapitulate the results published using the N2B27+2i media (von Meyenn *et al.*, 2016). Notwithstanding this, the transcriptomic profiles of N2B27 and KOSR cultured mESCs with 2i are reportedly broadly similar. A distinct class of genes found to be relatively more enriched in KOSR than N2B27 based media are associated with the 2-cell state, which includes Zscan4 for example. Therefore, KOSR+2i may model a broader spectrum of pluripotency (Hackett *et al.*, 2017). As outlined in the introduction, Zscan4 can negatively affect Dnmt1 and Uhrf1 levels and the regulation of DNA methylation by this route may be more prominent in KOSR+2i (Dan *et al.*, 2017). Other genes are also

differentially expressed between KOSR and N2B27 media and so future experiments may be best served by utilising both 2i compositions (Hackett *et al.*, 2017). Protein interactions that are common between both systems will be attractive for further characterisation. On the other hand, interactions that are found to be unique ought not to be disregarded for this reason. As genes are variably expressed throughout pre-implantation development, sequential mechanisms may contribute at different stages to promote or maintain the reprogramming of DNA methylation (Hackett *et al.*, 2017).

However, for both N2B27 and KOSR based systems with 2i, or indeed either of the Mek or Gsk3β inhibitors in isolation, global transcripts do not faithfully model earlier stages of development, such as E2.5 when DNA demethylation is also occurring (Hackett *et al.*, 2017). If a differential interacting partner found in the BioID screen is expressed in multiple stages of pre-implantation development, this would bolster the credibility that it may regulate Dnmt activity *in vivo* and warrant further investigation in the 2i systems. Confirming the expression profile of any candidates that may regulate DNA methylation and a successful functional study illustrating this role may render any questions regarding the importance of differences between KOSR and N2B27 2i media irrelevant.

In this present study, application of KOSR+2i was sufficient to drive DNA hypomethylation in mESCs. The rapid loss of the 5mC mark suggested an impairment in the maintenance machinery was largely responsible for this epigenetic reprogramming. In KOSR+2i, demethylation progressed at a similar rate to that reported in studies using N2B27+2i, wherein Dnmt1 activity was shown to be impaired through Uhrf1 downregulation. Furthermore, demethylation also occurred faster than in mESCs made deficient for the *de novo* Dnmts (Jackson *et al.*, 2004; von Meyenn *et al.*, 2016). Whilst Uhrf1 protein levels declined in KOSR+2i, this occurred after 2 days, by which time a 30% drop in 5mC was already recorded (see Figure 3.1). This suggested that alternative mechanisms are operative at earlier time points in the 2i transition that can account for the effects on DNA methylation. Using a proximity ligation method, BioID (Roux *et al.*, 2012), I sought to characterise these changes. By tagging both Uhrf1 and DNMT1 the differential regulation of the main components of the maintenance machinery can be investigated. A change in the composition of the complexes that either DNMT1 and Uhrf1 participates in within serum relative to the KOSR+2i condition, or vice versa, would present candidate regulators of these two proteins and consequently, the maintenance of DNA methylation.

A variety of proximity ligation techniques are now available in the biochemical toolbox to query novel interacting partners of a protein under investigation. BioID is highly relevant for the focus of this study as it is permissible to use harsh lysis conditions to liberate insoluble cellular fractions that are associated with chromatin (Varnaitė and MacNeill, 2016).

3.3.2 Optimisation of a BioID experiment to investigate the interactome of methylation maintenance machinery

To study the interactome of DNMT1 and Uhrf1 in mESCs, and the potential differential makeup of complexes under conditions promoting hyper- and hypomethylation, I have demonstrated that a number of suitable cell lines have been established and validated for use in BioID and pull-down experiments.

One potential drawback of the system I generated is the relatively low expression of Uhrf1-BirA*. Although this may be beneficial as it is also expressed at levels lower than endogenous Uhrf1, and therefore the potential for artefacts is minimised, this may exacerbate the issues related to the extensive biotinylation performed by the free BirA* control. As free BirA* displays a high degree of activity, there is the possibility that known and novel interactors of Uhrf1 are discarded as they are not enriched over this control.

The DNMT1-BirA* proteins are expressed to a similar extent and yet also have lower biotinylating activity than free BirA*. This effect may be expected based on the results in Figure 3.13 that show that there is a greater distribution of the free BirA* in cells and therefore the additional proteins biotinylated in Figures 3.9-3.12 may reside in the cytoplasm. This activity may also be an intrinsic property of conducting the experiment in mESCs however, a study into Autoimmune Regulator in the same line reported equivalent levels of activity were exhibited by a BirA* control tagged with mCherry and the fusion protein (Gu *et al.*, 2017). Others have also shown the utility in using a combination of BirA* fused with a nuclear localisation signal and a fluorescent tag as they yield different results (Lambert *et al.*, 2015).

At the expense of robustness, candidates enriched over the Empty vector control will therefore need to be included in the analysis. A more appropriate control to consider for similar experiments in the future is a line in which the protein of interest is overexpressed, such as DNMT1 and Uhrf1. This would recapitulate the potential changes in the proteome or the cell biology that may be induced by the overexpression of these epigenetic regulators as

BirA* fusions. Similarly, the BirA* fusion lines untreated with biotin, as published previously (Mulholland *et al.*, 2015), would also serve this purpose.

The limitations of BioID include those universal to all mass spectrometry methods, the most important factors being the efficiency with which lowly abundant proteins are detected and coverage. In order for proteins to be detected, they are required to be digested into peptides of a suitable length. However, this unavoidably increases the complexity of the mixture analysed and can give false positives (Han *et al.*, 2008). Other factors to consider that are unique to BioID include accessibility of the BirA* ligase to neighbouring proteins. If this is impaired by steric hindrance due to the position of BirA*, biotin ligation will occur less extensively and potential interactors or proximal proteins will not be identified. Furthermore, free biotin that has not be utilised by BirA* may compete with biotinylated proteins during the streptavidin pull-down. To minimise this, the wash buffer applied following the removal of media can be maintained on cells for an extended time to promote the diffusion of free biotin out of the cells, before commencing with the lysis procedure (Conlan *et al.*, 2018; Roux *et al.*, 2018).

In this study, the BioID screens will be complemented by conventional pull-down experiments. Whilst transient interactions will be difficult to maintain, relatively more stable interactions may be accounted for using this method. In contrast to BioID, which samples proteins in the vicinity, a pull-down may account for direct interactors or if it is preserved, components of an interacting complex. The candidates that fulfil the criteria of enrichment over the BirA* control and are detected in the pull-down will likely be the lead candidates. Initial validation steps will include performing endogenous IPs of Dnmt1 and Uhrf1 from serum and KOSR+2i maintained mESCs, followed by blotting for the candidates. Functional studies will follow in which the candidate is knocked down or overexpressed before the effect on DNA methylation is determined.

The BirA* of N-terminally tagged DNMT1 protein appears subject to cleavage. Figures 3.7 and 3.14. show free BirA* and intermediate sized proteins are present in the line. In the BioID screen, the free BirA* generated may therefore biotinylate proteins that are not proximal to DNMT1. When compared to the free BirA* control line however, these proteins may not be enriched and can therefore be accounted for, assuming the cleaved BirA* has a similar distribution. The efficiency of the Myc pull-down may also be compromised and it will

therefore be important to assess the enrichment of DNMT1 from the N-terminally tagged line relative to the control free BirA* line.

The use of BioID to assess the interactome of Dnmts has been published before (Mulholland *et al.*, 2015). In that particular study, Dnmt3b was selected and the differential interactions that occurred during epiblast differentiation were charted. The main focus however was to discuss the development of a system in which Dnmts, Uhrf1 and the Tet enzymes can be efficiently tagged with various cassettes using Bxb1-mediated recombination at a so-called multifunctional integrase (MIN) site introduced by CRISPR. The benefit of having an endogenously tagged Dnmt1 and Uhrf1 for example would be the absence of competition from an exogenously introduced fusion protein. This potential to compete for complexes may reduce the frequency of interactions with other proteins and therefore the detection over a background control is compromised. There is no guarantee however that the fusion protein will be expressed at a level comparable to the protein in a WT line. Another option available to negate the effects of competition would be to stably transfect into a null cell line. Various promoters could then be tested to attain a relevant level of protein expression. In both of these systems, the capability of the fusion protein to restore or maintain DNA methylation could then be assayed. If successful, this would provide further reassurance that localisation is equivalent to the endogenous protein and that interactions detected in the screen are reflective of events that truly occur *in vivo*.

However rather than optimising a system with BioID, there is a more efficient method now available that should be considered as the starting point in future experiments. This is known as TurboID and it represents the next generation in proximity labelling methods (Branon *et al.*, 2018). Developed from BirA* by directed evolution through serial rounds of error prone PCR and selection, the TurboID ligase is capable of attaining an equivalent level of biotinylation within 10 minutes than can be reached by BirA* in 18 hours. This has the potential to increase the depth of proteins captured that interact transiently, which may include more signalling mediators. A longer period of biotin incubation was shown to capture yet more proteins but this also correlated with an increased frequency of non-specific interactions. A free Turbo control may also reduce the background signal compared to free BirA* as vicinal proteins may not be saturated due to the requisite incubation time with biotin being considerably shorter (Branon *et al.*, 2018). BirA* has proven amenable to protein-fragment complementation (Trinkle-Mulcahy, 2019) and a similar development with TurboID

will be an exciting achievement. This could be exploited in a study like mine in which Dnmt1 and Uhrf1 fusions are expressed in the same line. As they are known interacting partners during the maintenance of DNA methylation, proteins would only be biotinylated when this complex is formed and as such the function of novel interactors could be inferred to assist in this role. Under conditions such as 2i, the expectation would be that a complex between Dnmt1 and Uhrf1 would be disrupted and therefore comparisons could be drawn regarding interactors in serum conditions.

Nevertheless, satisfied that the systems I generated had been optimised and appropriate conditions were available in which DNA methylation was modulated, I undertook a screen to identify novel interacting partners of DNMT1 and Uhrf1 in conditions that model naïve and committed states of pluripotency.

Chapter 4 – Proteomic screens to identify regulators of Uhrf1 and DNMT1 under conditions of DNA hypermethylation and hypomethylation

4.1 Introduction

Prior to the onset of this study, downregulation of Uhrf1 had not been identified as a mechanism contributing to the DNA hypomethylation observed in mESCs when transitioning from serum-based media to N2B27+2i (von Meyenn *et al.*, 2016). Nevertheless, results presented in Chapter 3 from an investigation into an alternative and understudied system, KOSR + 2i, suggests that mechanisms additional to Uhrf1 downregulation may also contribute at early stages to instigate DNA demethylation. In mESCs cultured in KOSR+2i medium, there is a 30% decline in methylation within 48 hours and this occurs prior to appreciable downregulation of Uhrf1 or Dnmt1 protein. Within this time window, I employed BioID to identify differential interacting partners of Uhrf1- and DNMT1-BirA* fusion proteins stably expressed in mESCs and cultured in KOSR+2i or serum-based media. This data will be complemented by a conventional Myc-tag mediated pull-down. These datasets combined will then influence the selection of candidate regulators of DNA methylation maintenance. Figure 4.1 represents an overview of the experimental scheme followed to obtain the data in this chapter.

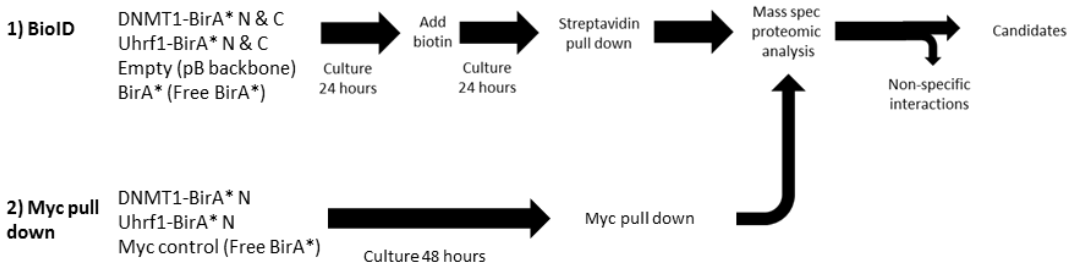


Figure 4.1 - Scheme of a BioID and pull-down screen to uncover the interacting partners of DNMT1 and Uhrf1 stably expressed in mESCs cultured in serum-based media or with KOSR + 2i. The definition of a candidate is led by the criteria of enrichment over at least one control cell line. In the BioID screen, a cell line expressing an empty vector (Empty) and another overexpressing BirA* (free BirA*) serve as controls. The BirA* constructs also contain a Myc-tag which can be exploited for pull-down. This dataset will strengthen the validity of the BioID data however, as this method is less suitable to detect transient interactions, it is not essential. For this set of experiments, non-specific interactions will be accounted for by the pull down of the free BirA*, referred to as the Myc pull-down control hereafter.

4.2 Results

Cell lysates for both BioID and Myc pull-down experiments were collected in triplicate for each cell line. The streptavidin-conjugated and Myc-trap beads used for BioID and the Myc-pull down respectively were then prepared following an appropriate washing procedure for mass spectrometry analyses. Further downstream processing and spectra analyses were performed by Jimi Wills (IGMM, University of Edinburgh) using MaxQuant software (Cox and Mann, 2008). Protein abundance was then calculated by myself from the relevant output file. As both screens were conducted without the use of isotopic labelling methods, protein quantification was measured in accordance with LFQ intensity values. This employs a number of algorithms for the purposes of normalising protein intensities and has been used for similar protein-protein interaction experiments with satisfactory results (Cox *et al.*, 2014).

4.2.1 Data processing

Statistical tests referred to throughout this chapter were two-tailed unpaired t-tests. These were applied to filter proteins that are significantly ($p \leq 0.05$) enriched over the Empty vector or free BirA* controls in serum-based media and KOSR+2i. For example, to isolate proteins enriched over the Empty vector control, the first array of data was the triplicate LFQ values for the Empty vector control and the second array the LFQ values of the N-terminally BirA* tagged Uhrf1 line for a particular protein. Therefore by comparing the mean LFQ values, those two-fold enriched in the experimental line relative to a control are not simply enriched because one replicate has a disproportionately high LFQ intensity value, but because all replicates have similar intensity values and are each enriched by two-fold over the control. The same process was applied to all of the samples and relevant controls, meaning the average LFQ intensity of a protein in the serum and KOSR+2i cultured experimental lines were compared relative to the serum and KOSR+2i cultured Empty, and separately, free BirA* line. With respect to the data from the Myc pull-down screen, LFQ intensities from the experimental sample were compared to values from the Myc pull-down control that was performed in the same experiment (see Figure 4.2).

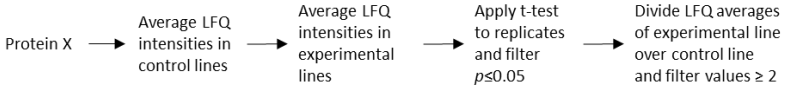


Figure 4.2 – Schematic of data analysis workflow. See main body of text for details.

The initial step of the data analysis was to assess the levels of Dnmt1 and Uhrf1 in the lines and experiments. As these are the 'baits' or fusion proteins that will biotinylate proteins, the intensity of these should be amongst the highest of the proteins detected if the pull-down had occurred as intended.

As shown in Figure 4.3, the levels of Uhrf1 in the BioID experiments are relatively consistent across the triplicates in the same culture condition. When applying a t-test as outlined above and calculating the fold-enrichment, N- and C- terminally tagged Uhrf1 are 198.73- and 123.54-fold more enriched over the average LFQ value calculated in the Empty vector control. Relative to BirA*, Uhrf1 is 2395.84- and 1489.33-fold more abundant, respectively. In KOSR+2i culture, these values are 2754.94 and 3644.88-fold for N-terminally tagged Uhrf1 over the Empty vector and BirA* controls and 2719.39 and 3597.85-fold for C-terminally tagged Uhrf1, respectively.

Instantly apparent in Figure 4.3 is the greater intensity of Uhrf1 in KOSR+2i compared to serum-based culture. This equates to a 2.66- and 4.22-fold difference between the media in the N and C-terminal tagged lines, respectively. In order to identify differential interactors, the LFQ values for proteins in the experimental lines maintained in serum-based media will require normalising through multiplication by this fold-difference. To account for the intensity that may be attributed to the binding of the non-biotinylated protein in the serum sample, an LFQ intensity that is 2-fold greater than the multiplication factor is regarded as differentially enriched. Therefore, proteins that are 5.32- and 8.44-fold more abundant in serum-based media by normalised LFQ value in the N- and C-terminally tagged lines are considered as differential interactors. On the other hand, an average LFQ intensity that is 2-fold greater for KOSR+2i compared to serum will be considered as the threshold above which a protein is differentially interacting.

The differential level of Uhrf1 is also observed in the results from the Myc pull-down screen. The KOSR+2i culture line exhibits 2.53-fold higher abundance of Uhrf1 on average when compared to the serum lines. As a consequence, serum LFQ intensities have also been normalised through multiplication by this factor and a 5.07-fold cut-off taken as the threshold above which a protein is considered to differentially interact in serum-based media. An enrichment in KOSR+2i that is 2-fold above the average serum value, as in the BioID experiment, has been taken as a differential interactor.

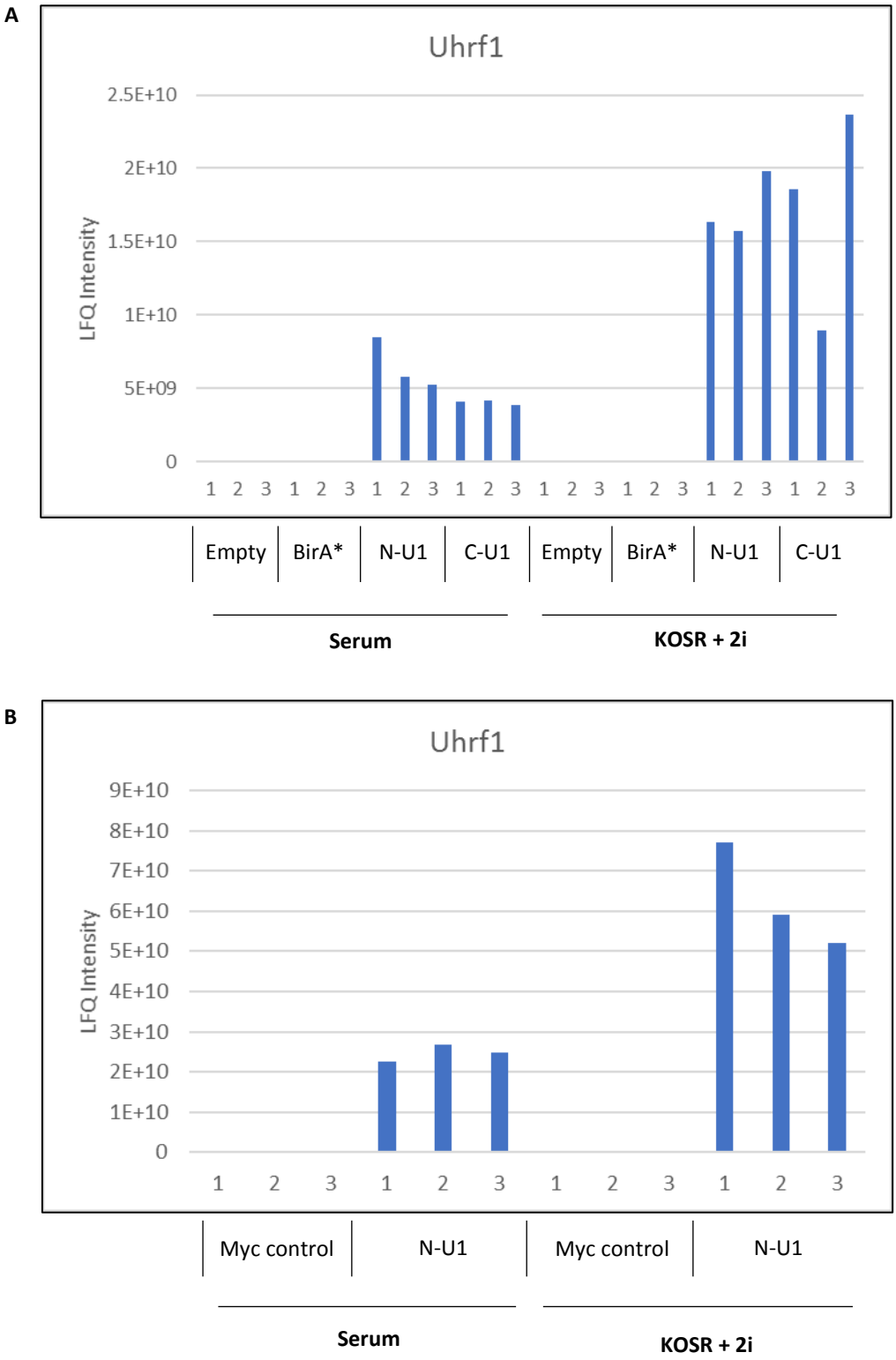


Figure 4.3 - *Figure legend on next page*

Figure on previous page **Figure 4.3** – LFQ Intensity of Uhrf1 calculated by mass spectrometry from **A.** BioID and **B.** Myc pull-down experiments in the cell lines and media conditions indicated. Numbers 1-3 under each bar represent the replicates of each group. Empty vector control (Empty), Myc control, BirA* N-terminally tagged Uhrf1 (N-U1), BirA* C-terminally tagged Uhrf1 (C-U1)

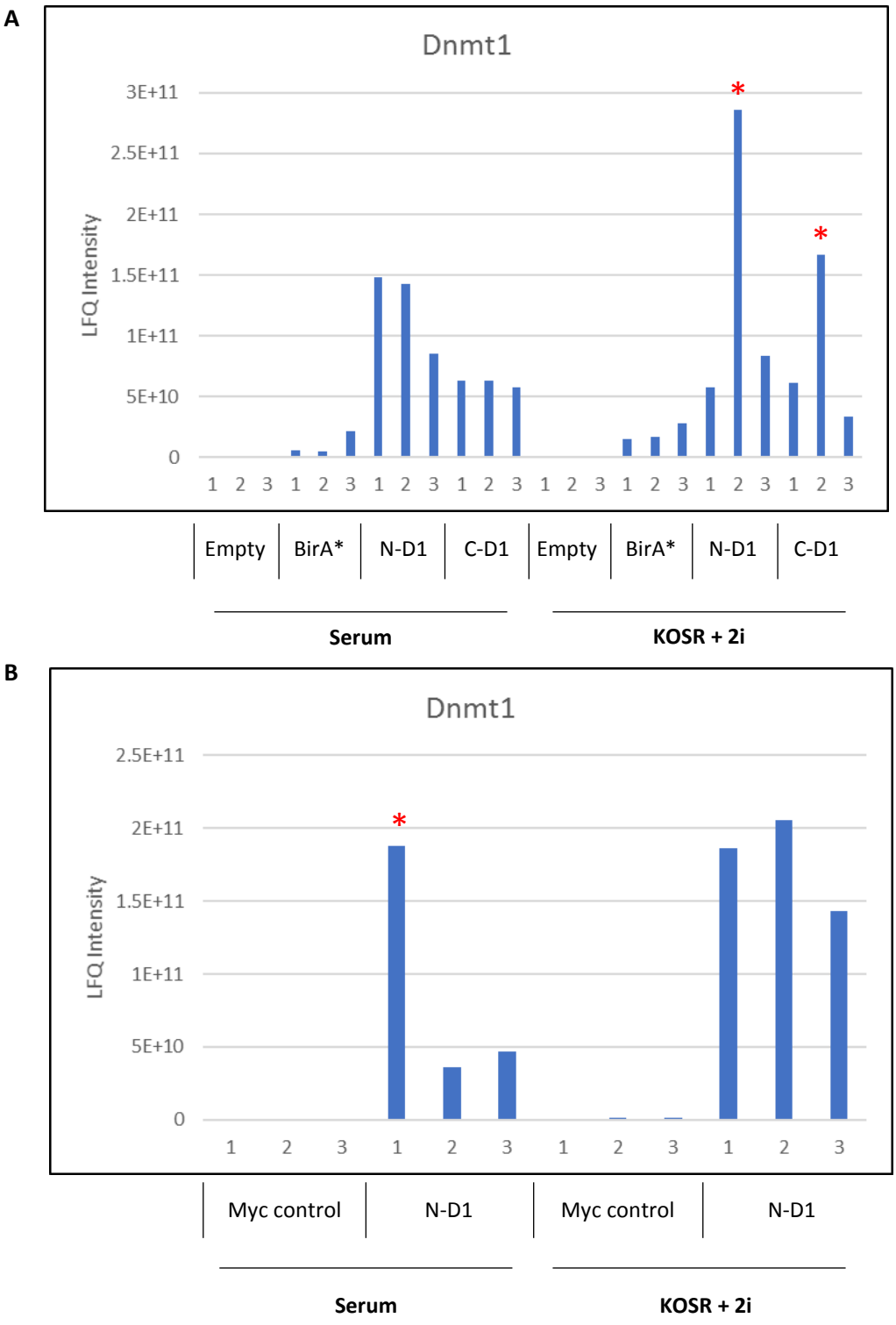


Figure 4.4 – *Figure legend on next page*

Figure on previous page **Figure 4.4** - LFQ Intensity of Dnmt1 calculated by mass spectrometry from **A.** BioID and **B.** Myc pull-down experiments in the cell lines and media conditions indicated. Numbers 1-3 under each bar represent the replicates of each group. Red asterixis denote the samples that were omitted from subsequent analyses. Empty vector control (Empty), Myc control, BirA* N-terminally tagged Dnmt1 (N-D1), BirA* C-terminally tagged Dnmt1 (C-D1)

Similar experiments with the DNMT1-BirA* lines were less successful. In the BioID screen, a replicate from each of the experimental lines cultured in KOSR+2i had to be omitted from the analysis owing to the large degree of variability. Furthermore, a replicate from the Myc pull-down in serum media has also been removed from consideration (Figure 4.4). With the inclusion of these replicates, a t-test comparing the average LFQ intensities in the experimental line and controls would otherwise have removed Dnmt1 itself from the list of enriched proteins, reflecting a statistically significant difference between the replicates. Therefore, rather than leading with this dataset, it has been used to supplement the findings from the Uhrf1 lines.

To facilitate the data analysis, a t-test was applied, as outlined above, to filter out proteins that are significantly (p>0.05) variable between the replicates. Those enriched two-fold over either the Empty or free BirA* controls were then isolated. With this criteria, 818 proteins were shown to be enriched over the Empty vector control in the N and C-terminally tagged lines collectively. Of this number, 498 were enriched in both the serum and KOSR+2i conditions whilst 66 and 254 specifically interacted, respectively. A considerably smaller number of 88 proteins were enriched over the free BirA* control. Of these, 26 were shared whilst 29 more were found uniquely in the serum condition and 33 in KOSR+2i (Figure 4.5).

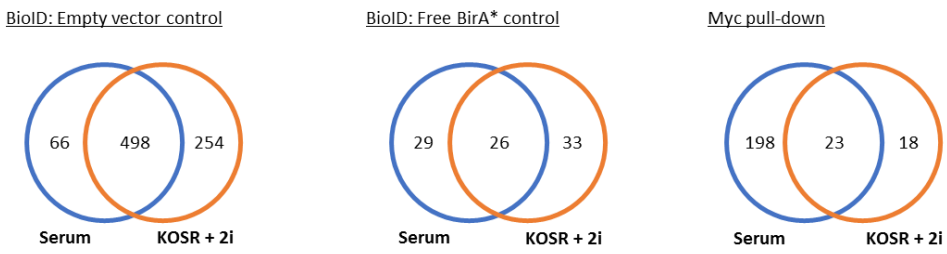


Figure 4.5 – The number of proteins from the Uhrf1 BioID and Myc pull-down screens enriched over the respective background controls (Empty vector and Free BirA*/Myc pull-down control) in serum and KOSR+2i media. Values that were at least 2-fold more abundant and not of a significantly different intensity between replicates were considered as enriched.

These lists were then ranked according to the average LFQ value in the appropriate experimental sample. The series of tables presented below show the fifty most abundant proteins that are enriched at least two-fold over the Empty vector controls. The number of proteins enriched over the free BirA* in some circumstances did not exceed fifty and they are presented in their entirety. Reassuringly, Uhrf1 was represented close to, if not the top, of all these lists (Table 4.1 A-H). Expanded tables detailing the number of peptides detected in each sample can be found in Appendix 1. Accompanying volcano plots generated by Jimi Wills (IGMM, University of Edinburgh) can be found in Appendix 2.

The accompanying Myc pull-down experiment yielded 239 proteins over the background control, 198 of which were uniquely enriched in serum, 18 in KOSR+2i and 23 were found in both conditions (Figure 4.5).

The data collected from the DNMT1-BirA* lines was also subject to the same processing methods. Underlying the deficiency of the dataset is the absence of DNMT1 by these measures from the proteins enriched over the BirA* control in the C-terminally tagged line cultured in KOSR+2i. Furthermore, the trend of increased bait levels with KOSR+2i culture compared to serum, as seen with the Uhrf1 lines, was observed in the Myc pull-down but not with BioID, suggesting cells were insufficiently lysed or beads were lost (Figure 4.4). Nevertheless, 475 shared proteins cleared the Empty vector background threshold and 239 and 256 were found specifically in serum and KOSR+2i conditions, respectively. As expected, proteins enriched over the BirA* control were fewer: 14 were found in both conditions and 20 each were unique in serum and KOSR+2i. Meanwhile, 15 and 43 proteins were unique in the serum and KOSR+2i Myc pull-down respectively and 10 were shared.

4.2.2 Visualising the data

STRING (https://string-db.org/) is an online database into which the names of individual proteins are inputted and a web of known or inferred interactors are returned. Interactions are depicted by a series of coloured lines, with each line representing a different method that was used to demonstrate or infer an interaction occurs. A key describing what the coloured lines represent is shown in Figure 4.6.

Similarly, a list of proteins can also be uploaded and databases are mined to ascertain if there are associations between them each. This is particularly useful in proteomic screens as it demonstrates whether detected proteins are likely to interact with the bait in a pull-down,

or the BirA* fusion protein in BioID, either directly or indirectly as they are a component of a larger interacting complex. Handily, functional enrichment analyses are also conducted which can be used to summarise the previously described role of the proteins inputted. Each of the tables (Table 4.1 A-H) presented were entered into STRING to assist in the visualisation of data and the corresponding graphics are shown below (Figure 4.7 A-H). Levels of confidence can be changed that may establish a relatively more tenuous association between proteins however, I have maintained the default 'medium' option throughout. The Gene Ontology terms associated with each list is also displayed as an inset.

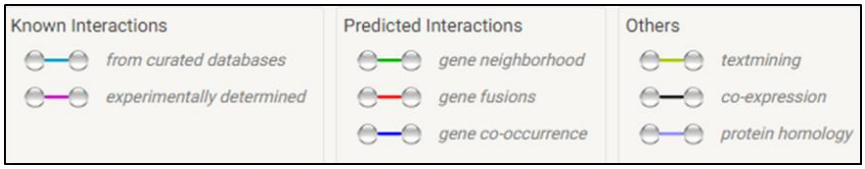


Figure 4.6 – Screenshot from the STRING database of the key for each coloured line that infers or denotes an interaction between two proteins.

Table 4.1 – The most abundant proteins ranked by average LFQ intensity detected in an N- and C-BirA* terminally tagged Uhrf1 line in a BioID experiment with mass spectrometry. Uhrf1 is highlighted in each table A-D) The fifty most abundant proteins enriched at least 2-fold over an Empty vector control. E-H) All proteins enriched at least two-fold over a free BirA* control.

A

N-terminus Uhrf1 Serum (enriched over Empty vector)			
Gene name	**Average LFQ**	**Fold enrichment**	***p* value**
Pcca	1.17E+10	2.35E+00	1.20E-02
Jmjd1c	7.28E+09	5.92E+01	2.40E-03
Acacb	6.80E+09	3.85E+00	2.39E-03
Uhrf1	6.50E+09	1.99E+02	3.08E-03
Mki67	5.13E+09	8.12E+03	5.10E-04
Hspa8	4.67E+09	3.50E+00	5.49E-03
Tpx2	3.63E+09	1.17E+03	6.99E-04
Rif1	3.09E+09	2.80E+02	4.10E-04
Smarca5	2.83E+09	3.24E+02	7.95E-04
Lig1	2.66E+09	1.90E+02	2.67E-03
Zfp62	2.59E+09	2.35E+02	1.13E-03
Nasp	2.49E+09	2.59E+03	1.10E-03
Numa1	2.44E+09	1.98E+02	1.11E-03
Trim28	2.13E+09	6.29E+02	3.23E-03
Nsd1	2.11E+09	7.75E+02	4.91E-04
Npm1	1.75E+09	4.21E+01	2.91E-03
Nup98	1.48E+09	1.96E+00	1.72E-02
Chd4	1.40E+09	3.38E+02	1.05E-03
Lbr	1.18E+09	1.55E+01	8.54E-05
Ncl	1.07E+09	7.58E+01	3.11E-03
Msh6	9.41E+08	6.54E+01	1.35E-03
Tet1	8.02E+08	2.15E+03	7.03E-04
Ppp1r10	7.71E+08	7.71E+08	1.75E-03
Wdhd1	6.74E+08	6.74E+08	1.08E-03
Top2a	6.56E+08	2.12E+01	2.64E-03
Brd4	5.95E+08	5.95E+08	1.16E-03
Hist1/2/3	5.90E+08	1.09E+01	8.25E-05
Usp7	5.72E+08	4.67E+01	4.40E-03
Sgol2	5.62E+08	9.90E+00	1.99E-03
Hist1h4a	5.56E+08	1.16E+01	3.06E-03
Ahctf1	5.37E+08	4.15E+01	1.01E-03
Hspa5	5.28E+08	3.99E+00	7.56E-03
Tmpo	5.14E+08	1.76E+01	2.63E-02
U2surp	4.91E+08	2.43E+00	1.29E-02
Ddx5	4.70E+08	2.15E+00	3.43E-02
Set	4.55E+08	4.55E+08	4.36E-04
Dhx9	4.42E+08	2.77E+00	1.64E-03
Znf281	4.41E+08	1.67E+01	1.52E-02
Ranbp2	4.33E+08	5.57E+00	5.96E-03
Atrx	4.25E+08	1.45E+03	2.19E-03
Dnmt1	4.05E+08	1.29E+01	4.28E-04
Igf2bp1	3.79E+08	2.68E+00	2.87E-02
Msh2	3.63E+08	2.09E+00	3.05E-02
Fam208a	3.62E+08	1.66E+01	1.67E-03
Hcfc1	3.54E+08	1.88E+03	4.38E-03
Hnrnpu	3.50E+08	2.94E+00	4.53E-03
Chd8	3.46E+08	4.57E+02	4.90E-04
Ddx46	3.46E+08	3.01E+02	2.90E-03
Trip12	3.38E+08	5.66E+00	3.77E-04
Crk	3.32E+08	5.53E+02	1.13E-03

B

N-terminus Uhrf1 KOSR + 2i (enriched over Empty vector)			
Gene name	**Average LFQ**	**Fold enrichment**	***p* value**
Uhrf1	1.73E+10	2.75E+03	1.77E-04
Pcca	9.74E+09	2.94E+00	7.55E-04
Tpx2	6.05E+09	1.88E+03	3.25E-07
Hspa8	5.36E+09	2.56E+00	2.53E-04
Acacb	5.04E+09	4.38E+00	4.75E-05
Lig1	4.90E+09	1.80E+02	1.15E-09
Jmjd1c	4.71E+09	6.87E+01	2.70E-05
Sall4	4.63E+09	1.30E+03	4.00E-02
Mki67	4.00E+09	1.15E+04	1.30E-02
Smarca5	3.98E+09	3.08E+02	6.82E-04
Nasp	2.94E+09	1.11E+03	2.65E-05
Trim28	2.74E+09	6.64E+02	2.26E-06
Npm1	2.69E+09	1.67E+02	4.60E-06
Mospd3	2.68E+09	2.68E+09	4.86E-02
Nsd1	2.37E+09	9.67E+03	1.01E-04
Zfp62	2.28E+09	1.76E+02	1.50E-05
Lbr	1.87E+09	2.61E+01	1.53E-05
Chd4	1.68E+09	4.95E+02	1.81E-04
Rif1	1.59E+09	2.37E+02	3.93E-04
Msh6	1.52E+09	5.29E+01	1.80E-08
Numa1	1.40E+09	5.28E+01	4.85E-04
Ncl	1.39E+09	2.04E+02	4.90E-04
Usp7	1.05E+09	1.17E+02	1.02E-05
Ppp1r10	9.84E+08	9.84E+08	1.16E-04
Wdhd1	9.33E+08	9.33E+08	9.32E-08
Set	9.28E+08	9.28E+08	2.74E-05
Tmpo	8.81E+08	6.26E+02	9.12E-03
Kdm3b	8.65E+08	3.22E+01	1.71E-05
Hist1/2/3	8.45E+08	2.00E+01	4.71E-05
Top2a	8.45E+08	2.63E+01	1.11E-04
Hist1h4a	8.20E+08	3.11E+01	2.76E-04
Brd4	7.97E+08	7.97E+08	5.51E-04
Atrx	7.54E+08	7.67E+02	1.04E-02
Tet1	7.50E+08	7.50E+08	3.71E-04
Hspa5	7.33E+08	3.41E+00	8.47E-04
Ddx46	7.30E+08	1.55E+02	6.26E-06
Txnl1	6.79E+08	3.68E+01	4.49E-05
Sgol2	6.45E+08	1.03E+01	2.48E-03
Crk	6.42E+08	6.42E+08	9.12E-05
U2surp	6.20E+08	2.37E+00	1.14E-04
Znf281	6.17E+08	6.03E+02	7.14E-03
Ddx5	6.07E+08	2.46E+00	1.05E-04
Ranbp2	6.07E+08	8.70E+00	2.02E-05
Kif4	5.98E+08	1.32E+01	4.78E-04
Taf7	5.97E+08	6.16E+03	9.60E-04
Trip12	5.94E+08	7.82E+00	6.87E-04
Sart3	5.94E+08	1.02E+03	1.01E-07
Hells	5.76E+08	1.73E+01	1.25E-06
Wapal	5.68E+08	1.80E+02	5.94E-06
Ahctf1	5.67E+08	4.26E+01	3.77E-04

C

C-terminus Uhrf1 Serum (enriched over Empty vector)			
Gene name	**Average LFQ**	**Fold enrichment**	**p value**
Acacb	4.82E+09	2.73E+00	4.51E-05
Uhrf1	4.04E+09	1.24E+02	2.74E-06
Jmjd1c	3.38E+09	2.74E+01	2.36E-05
Hspa8	2.76E+09	2.07E+00	2.00E-03
Mki67	2.28E+09	3.60E+03	1.05E-02
Numa1	2.08E+09	1.69E+02	7.19E-04
Zfp62	1.86E+09	1.69E+02	4.37E-06
Sall4	1.81E+09	3.98E+02	3.69E-06
Tpx2	1.65E+09	5.33E+02	4.50E-03
Smarca5	1.43E+09	1.63E+02	1.06E-05
Rif1	1.23E+09	1.12E+02	1.39E-04
Nsd1	1.02E+09	3.74E+02	3.83E-05
Lig1	9.05E+08	6.46E+01	3.46E-05
Trim28	7.55E+08	2.22E+02	2.60E-05
Npm1	6.91E+08	1.66E+01	3.99E-05
Mospd3	6.81E+08	6.81E+08	2.13E-03
Nasp	6.62E+08	6.88E+02	1.30E-03
Ncl	5.95E+08	4.20E+01	1.32E-04
Chd4	5.48E+08	1.32E+02	3.41E-06
Ppp1r10	4.97E+08	4.97E+08	9.85E-05
Lbr	4.59E+08	6.03E+00	5.61E-05
Msh6	4.40E+08	3.05E+01	7.95E-06
Top2a	4.11E+08	1.33E+01	6.71E-06
Tet1	3.67E+08	9.82E+02	2.67E-04
Wdhd1	3.03E+08	3.03E+08	9.06E-06
Hist1h4a	2.92E+08	6.07E+00	2.76E-04
Ranbp2	2.91E+08	3.74E+00	6.76E-04
Tmpo	2.77E+08	9.50E+00	8.25E-04
Hspa5	2.61E+08	1.97E+00	5.09E-03
Sgol2	2.59E+08	4.56E+00	3.27E-02
Hist1h/2h	2.56E+08	4.73E+00	6.94E-05
Ahctf1	2.45E+08	1.89E+01	5.06E-04
Trip12	2.32E+08	3.87E+00	5.61E-04
Atrx	2.10E+08	7.15E+02	7.63E-04
Dnmt1	2.06E+08	6.54E+00	1.60E-04
Set	2.01E+08	2.01E+08	7.20E-05
Znf281	1.99E+08	7.55E+00	1.57E-02
Brd4	1.93E+08	1.93E+08	1.56E-04
Kif4	1.66E+08	1.66E+08	8.48E-04
Bptf	1.59E+08	1.59E+08	4.91E-04
Hist1h3a/b/H3f3a/c	1.50E+08	3.94E+00	2.05E-03
Mre11a	1.48E+08	1.48E+08	1.32E-04
Tox4	1.47E+08	1.47E+08	4.74E-05
Pcnp	1.37E+08	1.37E+08	3.57E-05
Chd8	1.35E+08	1.79E+02	7.98E-04
Orc2	1.32E+08	1.32E+08	2.31E-04
Anln	1.31E+08	1.31E+08	3.06E-05
Hcfc1	1.26E+08	6.68E+02	1.26E-03
Kdm3b	1.19E+08	1.03E+01	9.38E-04
Mga	1.19E+08	1.19E+08	3.23E-06

D

C-terminus Uhrf1 KOSR + 2i (enriched over Empty vector)			
Gene names	**Average LFQ**	**Fold enrichment**	**p value**
Uhrf1	1.71E+10	2.72E+03	1.66E-02
Pcca	9.04E+09	2.73E+00	2.54E-02
Sall4	5.35E+09	1.51E+03	1.47E-02
Tpx2	5.14E+09	1.59E+03	6.92E-03
Acacb	4.26E+09	3.70E+00	8.19E-03
Jmjd1c	3.39E+09	4.95E+01	9.21E-03
Smarca5	3.21E+09	2.49E+02	9.19E-03
Zfp62	2.63E+09	2.04E+02	1.36E-02
Lig1	2.32E+09	8.50E+01	4.03E-03
Mki67	2.24E+09	6.43E+03	3.88E-04
Numa1	1.86E+09	6.99E+01	4.95E-03
Npm1	1.75E+09	1.08E+02	7.02E-04
Nsd1	1.51E+09	6.18E+03	1.52E-02
Trim28	1.35E+09	3.26E+02	3.15E-03
Nasp	1.29E+09	4.86E+02	2.86E-03
Ncl	1.23E+09	1.80E+02	9.43E-04
Msh6	1.12E+09	3.89E+01	6.71E-03
Chd4	1.10E+09	3.23E+02	9.14E-03
Ppp1r10	9.79E+08	9.79E+08	5.28E-03
Lbr	9.71E+08	1.35E+01	6.44E-03
Rif1	9.44E+08	1.41E+02	4.98E-03
Trip12	6.65E+08	8.76E+00	1.24E-02
Tmpo	6.41E+08	4.55E+02	1.90E-03
Ranbp2	6.16E+08	8.83E+00	8.27E-03
Set	6.06E+08	6.06E+08	2.93E-03
Hist1h4a	5.92E+08	2.24E+01	1.62E-02
Wdhd1	5.91E+08	5.91E+08	2.00E-03
Hspa5	5.81E+08	2.70E+00	1.71E-02
Kdm3b	5.77E+08	2.15E+01	5.70E-03
Sgol2	5.67E+08	9.04E+00	1.41E-02
Top2a	5.50E+08	1.71E+01	1.06E-03
Atrx	5.42E+08	5.51E+02	4.53E-03
Brd4	5.07E+08	5.07E+08	7.27E-03
Znf281	5.06E+08	4.95E+02	1.71E-02
Hist1h/2h	4.77E+08	1.13E+01	2.29E-03
Dnmt1	4.77E+08	2.82E+01	1.74E-02
Taf7	4.75E+08	4.90E+03	1.96E-03
Kif4	4.58E+08	1.01E+01	1.43E-02
Ahctf1	4.51E+08	3.39E+01	4.28E-03
Tet1	4.46E+08	4.46E+08	7.12E-03
Hnrnpu	4.08E+08	3.35E+00	1.81E-02
Ddx46	3.96E+08	8.42E+01	5.83E-03
Hells	3.91E+08	1.17E+01	5.18E-03
Pcnp	3.85E+08	1.84E+03	2.06E-02
Hist1h2/H2	3.80E+08	5.87E+00	2.17E-02
Crk	3.53E+08	3.53E+08	3.08E-03
Igf2bp1	3.52E+08	2.91E+00	1.12E-02
Tox4	3.41E+08	3.41E+08	1.39E-02
Anln	3.40E+08	4.31E+02	2.21E-03
Chd8	3.22E+08	1.10E+03	2.89E-03

E

N-terminus Uhrf1 Serum (enriched over BirA*)			
Gene names	Average LFQ	Fold enrichment	p value
Acaca	8.43E+10	4.16E+00	2.51E-03
Mccc1	1.62E+10	2.60E+00	3.56E-03
Pcca	1.17E+10	2.50E+00	9.70E-03
Acacb	6.80E+09	8.90E+00	1.16E-03
Uhrf1	6.50E+09	2.40E+03	3.03E-03
Mki67	5.13E+09	2.44E+00	7.44E-03
Dnajb6	2.91E+08	2.08E+00	2.46E-03
Mtco2	1.11E+08	3.68E+00	2.21E-02
Utf1	8.75E+07	6.07E+00	1.26E-02
Tubg1	5.78E+07	2.72E+00	4.09E-02
Bard1	4.32E+07	2.34E+00	5.24E-03
Apobec3	3.94E+07	3.94E+07	1.49E-03
Pccb	3.79E+07	3.79E+07	4.91E-03
Derl1	3.76E+07	5.29E+00	3.90E-02
Cul7	3.33E+07	3.33E+07	3.19E-05
Zfa/x	2.95E+07	5.89E+00	1.93E-02
Vdac3	2.48E+07	3.89E+00	4.78E-02
Gldc	2.32E+07	4.26E+00	3.32E-02
Ube2n	2.29E+07	2.29E+07	5.56E-06
Dppa3	1.82E+07	1.82E+07	1.34E-03
Nsun5	1.79E+07	1.79E+07	3.27E-04
Dimt1	1.71E+07	5.82E+00	1.36E-02
Hist1h1b	1.64E+07	1.64E+07	1.44E-03
Timm50	1.59E+07	6.62E+00	6.82E-03
Pcid2	1.57E+07	1.57E+07	1.08E-02
Papd4	1.50E+07	5.44E+00	1.46E-02
Cdk5	1.42E+07	1.42E+07	2.54E-05
Tfb1m	1.39E+07	1.39E+07	7.18E-03
Znf326	1.31E+07	1.31E+07	2.16E-03
Pdk1	1.28E+07	1.28E+07	1.23E-04
Hmg20a	1.25E+07	1.25E+07	2.00E-02
AHSG	1.11E+07	1.11E+07	1.15E-03
Slc25a13	1.03E+07	1.03E+07	5.63E-04
Gcat	9.47E+06	9.47E+06	4.13E-02
Riok2	8.73E+06	8.73E+06	1.04E-03
Nmd3	8.50E+06	8.50E+06	3.59E-05
Bcl9l	7.26E+06	7.26E+06	3.80E-03
Ercc4	6.46E+06	4.60E+00	3.10E-02
Slc25a1	6.05E+06	6.05E+06	2.45E-03
Eppk1	5.52E+06	5.52E+06	3.49E-04
Idh3b	4.71E+06	4.71E+06	6.88E-04
Ddx47	4.71E+06	4.71E+06	5.49E-03
Usp29	4.10E+06	4.10E+06	1.94E-03
Anp32a/b/c	3.86E+06	3.86E+06	1.67E-03

F

N-terminus Uhrf1 KOSR +2i (enriched over BirA*)			
Gene names	Average LFQ	Fold enrichment	p value
Acaca	5.52E+10	3.30E+00	5.70E-04
Uhrf1	1.73E+10	3.64E+03	1.77E-04
Mccc1	1.48E+10	2.72E+00	2.05E-04
Pcca	9.74E+09	3.27E+00	1.55E-04
Acacb	5.04E+09	8.87E+00	4.59E-07
Mki67	4.00E+09	3.81E+00	3.49E-02
Numa1	1.40E+09	1.99E+00	7.60E-03
Tet1	7.50E+08	3.44E+00	1.55E-03
Txnl1	6.79E+08	1.97E+00	9.55E-04
Dnajb6	3.24E+08	2.30E+00	1.12E-04
Dppa3	1.91E+08	1.91E+08	1.19E-03
Rere	1.88E+08	5.79E+00	1.02E-03
Inpp4a	1.15E+08	4.84E+00	2.20E-02
Bard1	1.08E+08	3.06E+00	1.69E-04
Kat6b	6.62E+07	2.33E+00	1.53E-02
Auts2	5.26E+07	1.01E+01	1.21E-04
Smchd1	5.25E+07	4.93E+00	7.98E-03
Bckdk	3.90E+07	2.58E+00	4.98E-02
Utf1	3.81E+07	3.81E+07	4.20E-04
Zfa/x	3.69E+07	3.69E+07	1.78E-03
Baz2b	3.63E+07	5.94E+00	8.62E-03
Pccb	3.29E+07	3.29E+07	1.40E-03
Ube2n	2.50E+07	3.90E+00	4.60E-02
Ywhah	2.37E+07	2.37E+07	8.15E-03
Wiz	2.14E+07	4.23E+00	3.53E-02
Fads1	1.95E+07	1.95E+07	6.59E-04
Srcap	1.72E+07	1.72E+07	3.14E-04
Nsun5	1.68E+07	1.68E+07	9.97E-06
Pcid2	1.46E+07	1.46E+07	4.98E-04
Anp32a/b/c	1.37E+07	1.37E+07	1.79E-02
Zkscan3	1.36E+07	1.36E+07	7.51E-03
Rad21	1.13E+07	1.13E+07	1.73E-04
Ercc4	1.07E+07	1.07E+07	1.32E-03
Riok2	9.43E+06	9.43E+06	5.10E-05
Serpinh1	8.82E+06	8.82E+06	2.78E-03
Chtf8	8.74E+06	8.74E+06	2.33E-06
Ino80	8.69E+06	8.69E+06	4.28E-03
Mccc2	8.52E+06	1.24E+01	1.91E-02
Smc6	8.40E+06	8.40E+06	1.78E-03
Rnf216	7.53E+06	7.53E+06	7.55E-04
Nmd3	6.47E+06	6.47E+06	1.07E-04
Hsd17b12	6.21E+06	6.21E+06	1.45E-04
Hspa14	6.09E+06	6.09E+06	7.09E-04
Idh3b	5.27E+06	5.27E+06	8.22E-03
Dnmt3a	4.87E+06	4.87E+06	5.63E-04

G

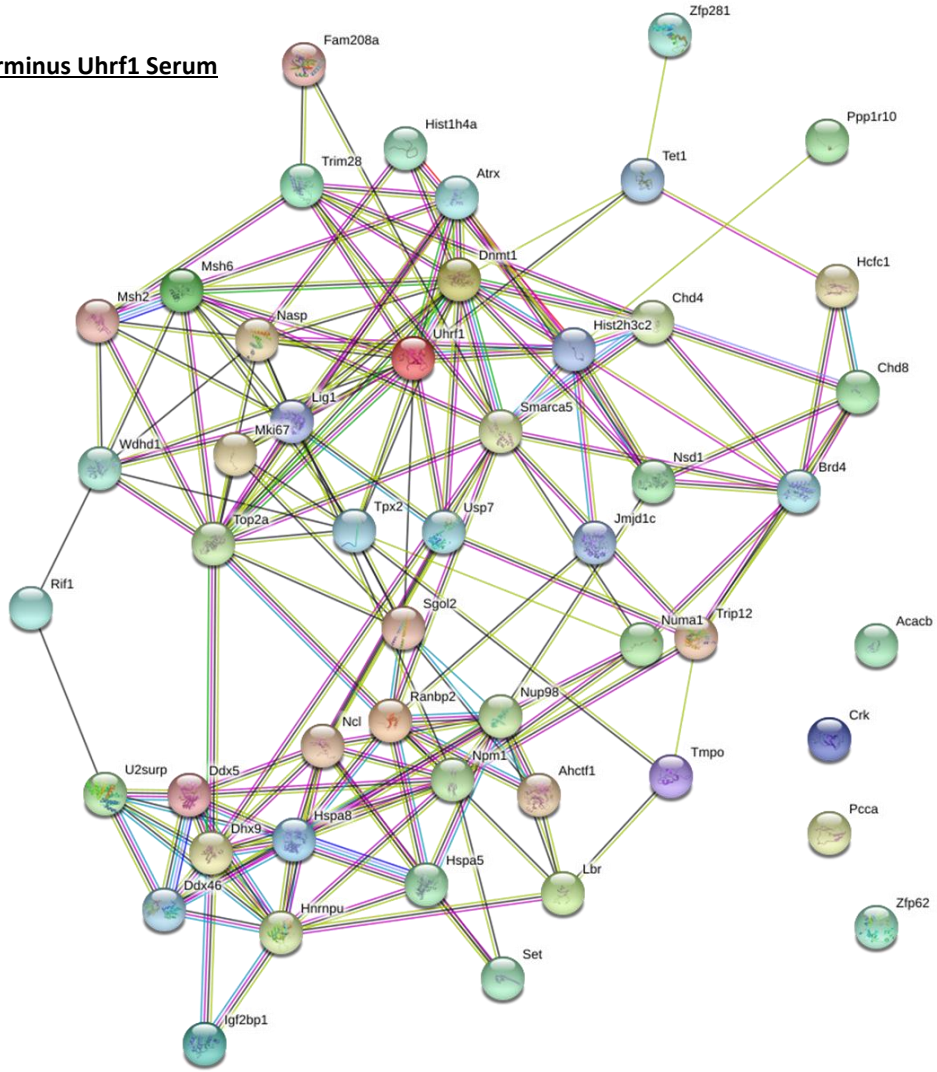
C-terminus Uhrf1 Serum (enriched over BirA*)			
Gene names	**Average LFQ**	**Fold enrichment**	**p value**
Acaca	5.53E+10	2.73E+00	5.40E-05
Mccc1	1.23E+10	1.97E+00	7.85E-04
Acacb	4.82E+09	6.32E+00	4.59E-06
Uhrf1	4.04E+09	1.49E+03	2.58E-06
Rere	1.08E+08	2.73E+00	1.07E-02
Utf1	8.57E+07	5.95E+00	1.04E-02
Pccb	3.32E+07	3.32E+07	2.85E-02
Cul7	2.57E+07	2.57E+07	2.54E-03
Apobec3	2.50E+07	2.50E+07	1.58E-03
Dppa3	1.87E+07	1.87E+07	4.44E-02
Mccc2	1.70E+07	1.70E+07	3.75E-02
Papd4	1.22E+07	4.42E+00	2.95E-02
Hist1h1b	1.12E+07	1.12E+07	2.37E-03
Nsun5	1.00E+07	1.00E+07	2.00E-07
Timm50	9.90E+06	4.13E+00	3.82E-02
Zbtb48	8.92E+06	8.92E+06	2.77E-02
Pdk1	8.39E+06	8.39E+06	3.40E-03
Hmg20a	8.38E+06	8.38E+06	8.04E-05
Cdk5	7.77E+06	7.77E+06	2.75E-04
Nup43	7.60E+06	7.60E+06	1.38E-03
Slc25a13	7.10E+06	7.10E+06	1.93E-03
Riok2	7.07E+06	7.07E+06	2.52E-03
Nmd3	5.66E+06	5.66E+06	1.43E-04
Bcl9l	4.98E+06	4.98E+06	8.38E-04
Gli3	4.87E+06	4.87E+06	3.04E-03
Slc25a1	4.83E+06	4.83E+06	3.89E-03
Gnl3	3.98E+06	3.98E+06	2.21E-03
Nt5dc2	3.54E+06	3.54E+06	1.24E-02
Tpi1	1.88E+06	1.88E+06	1.94E-04

H

C-terminus Uhrf1 KOSR + 2i (enriched over BirA*)			
Gene names	**Average LFQ**	**Fold enrichment**	**p value**
Acaca	4.42E+10	2.64E+00	3.35E-02
Uhrf1	1.71E+10	3.60E+03	1.66E-02
Mccc1	1.38E+10	2.55E+00	2.43E-02
Pcca	9.04E+09	3.03E+00	1.82E-02
Acacb	4.26E+09	7.49E+00	3.71E-03
Mki67	2.24E+09	2.13E+00	5.17E-03
Numa1	1.86E+09	2.64E+00	2.48E-02
Dnajb6	3.58E+08	2.53E+00	2.77E-02
Dppa3	1.95E+08	1.95E+08	4.75E-04
Rere	1.79E+08	5.52E+00	6.35E-03
Auts2	5.38E+07	1.03E+01	3.43E-03
Utf1	4.29E+07	4.29E+07	7.34E-03
Pccb	3.48E+07	3.48E+07	1.55E-02
Baz2b	3.40E+07	5.57E+00	4.32E-02
Wiz	2.43E+07	4.81E+00	3.40E-02
Fads1	1.70E+07	1.70E+07	3.20E-03
Dnajb2	1.43E+07	1.43E+07	9.35E-03
Nsun5	1.36E+07	1.36E+07	4.12E-03
Ywhah	1.34E+07	1.34E+07	9.83E-03
Fbrsl1	1.29E+07	1.29E+07	5.42E-03
Zfa/x	1.11E+07	1.11E+07	1.09E-02
Srcap	1.03E+07	1.03E+07	1.89E-03
Chtf8	1.03E+07	1.03E+07	7.07E-09
Huwe1	1.02E+07	1.02E+07	1.92E-03
Csnk1e	9.16E+06	9.16E+06	1.54E-04
Riok2	8.76E+06	8.76E+06	1.96E-02
Zfp563	8.70E+06	8.70E+06	8.67E-03
Rad21	8.25E+06	8.25E+06	6.48E-03
Nup43	7.97E+06	7.97E+06	1.69E-03
Gcat	6.56E+06	6.56E+06	1.16E-03
Rnf216	6.14E+06	6.14E+06	1.69E-03
Nmd3	5.87E+06	5.87E+06	1.85E-03
Serpinh1	5.67E+06	5.67E+06	1.09E-04
Slc25a11	5.35E+06	5.35E+06	7.62E-04
Sec24c	5.00E+06	5.00E+06	1.76E-03
Anp32a/b/c	4.77E+06	4.77E+06	8.90E-03

A

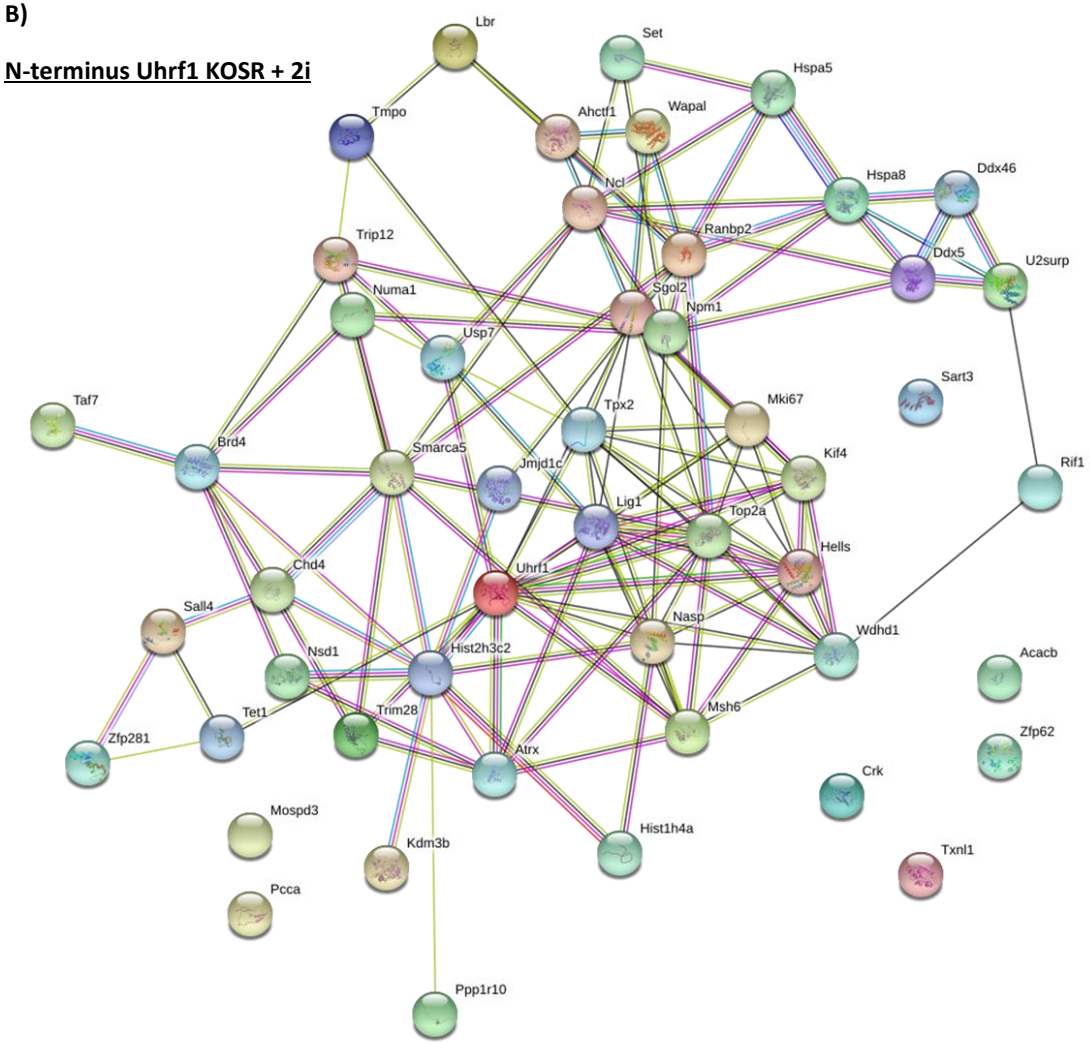
N-terminus Uhrf1 Serum



Biological Process (GO)		
Pathway Description	**Count in gene set**	**False discovery rate**
Chromosome organisation	26	5.45×10^{-20}
Chromatin organisation	20	1.02×10^{-15}
Regulation of nucleobase-containing compound metabolic process	35	5.38×10^{-15}
Nucleic acid metabolic process	33	6.03×10^{-14}
Regulation of chromosome organisation	15	1.53×10^{-13}

B)

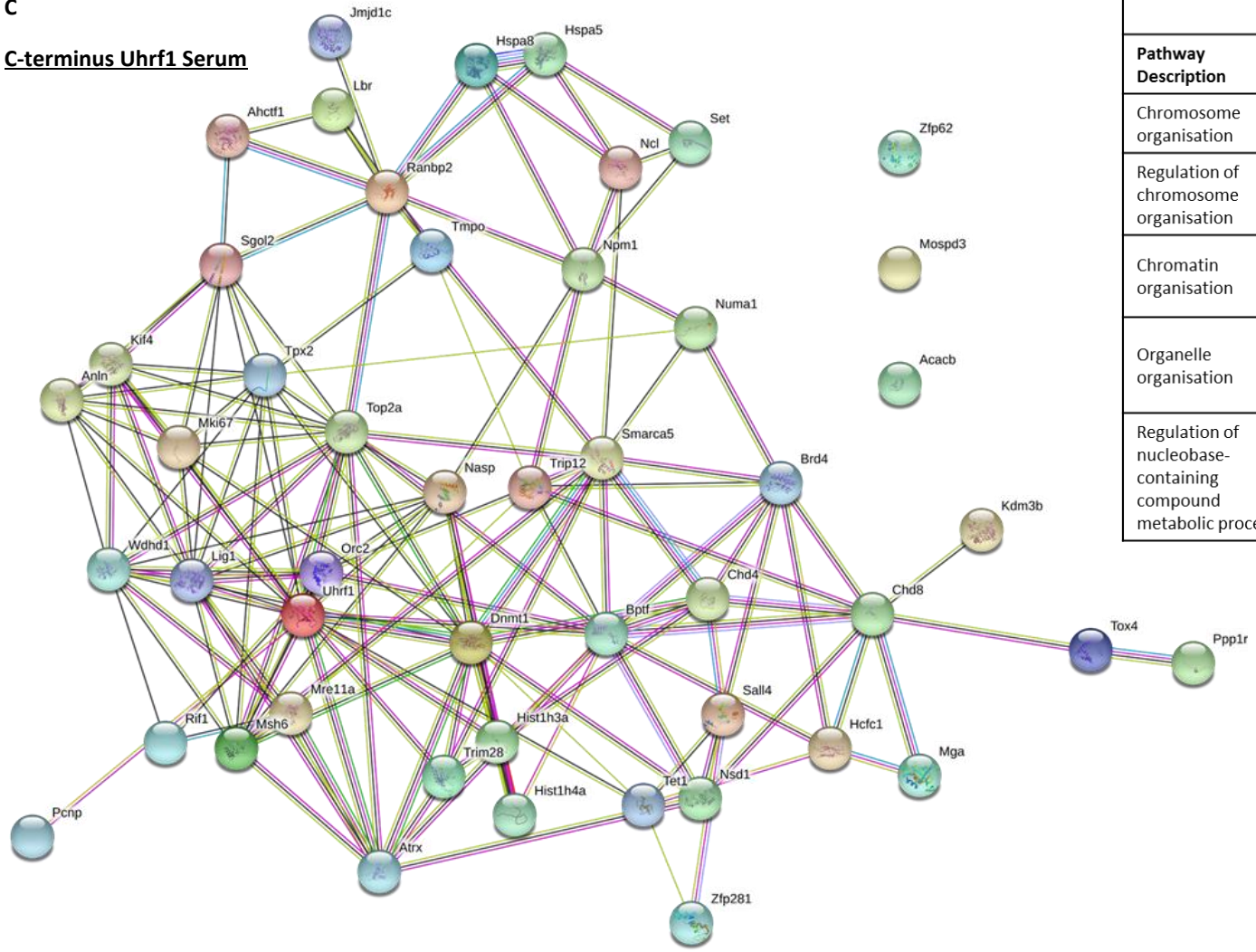
N-terminus Uhrf1 KOSR + 2i



Biological Process (GO)		
Pathway Description	**Count in gene set**	**False discovery rate**
Chromosome organisation	24	2.74×10^{-17}
Regulation of chromosome organisation	16	1.17×10^{-14}
Chromatin organisation	19	1.41×10^{-14}
Organelle organisation	29	2.07×10^{-11}
Regulation of nucleobase-containing metabolic processes	31	3.58×10^{-11}

C

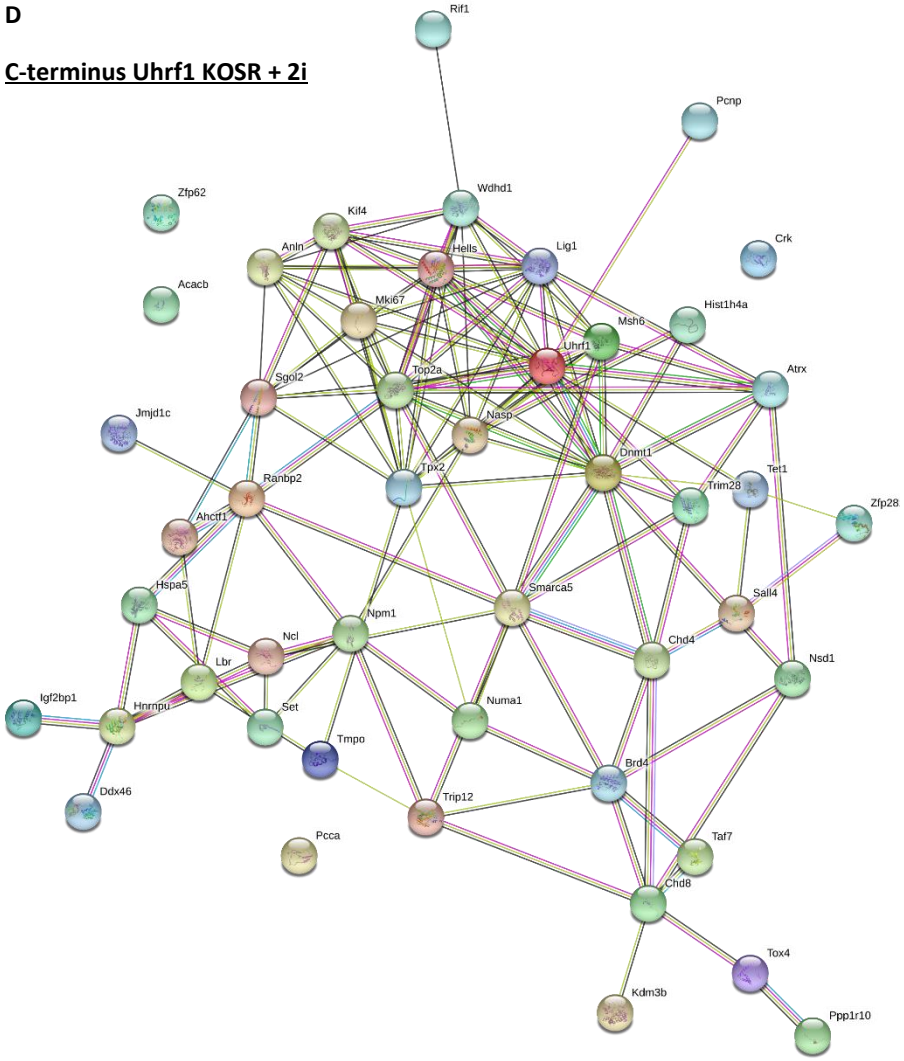
C-terminus Uhrf1 Serum



Biological Process (GO)		
Pathway Description	**Count in gene set**	**False discovery rate**
Chromosome organisation	23	3.07×10^{-16}
Regulation of chromosome organisation	15	2.38×10^{-13}
Chromatin organisation	18	2.38×10^{-13}
Organelle organisation	28	8.99×10^{-11}
Regulation of nucleobase-containing compound metabolic process	30	1.37×10^{-10}

D

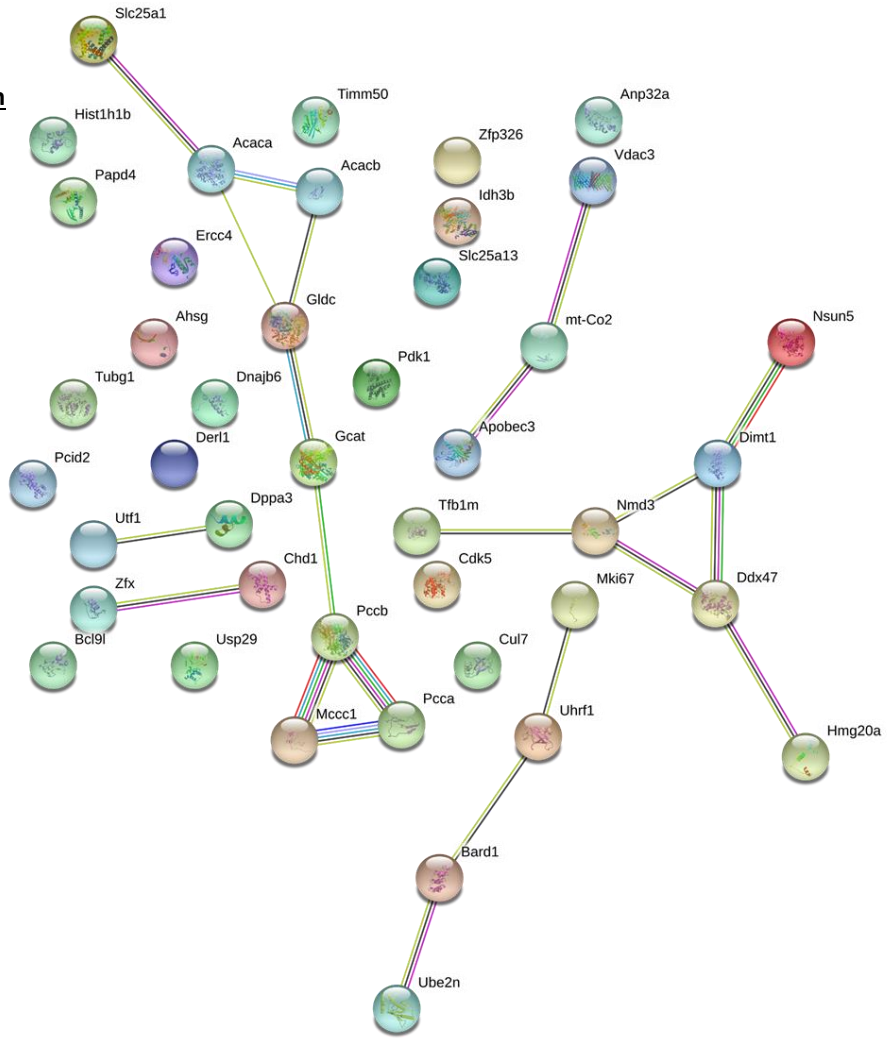
C-terminus Uhrf1 KOSR + 2i



Biological Process (GO)		
Pathway Description	**Count in gene set**	**False discovery rate**
Chromosome organisation	23	1.75×10^{-16}
Regulation of chromosome organisation	16	5.44×10^{-15}
Chromatin organisation	19	5.58×10^{-15}
Regulation of organelle organisation	20	2.84×10^{-11}
Regulation of chromatin organisation	11	4.80×10^{-11}

E

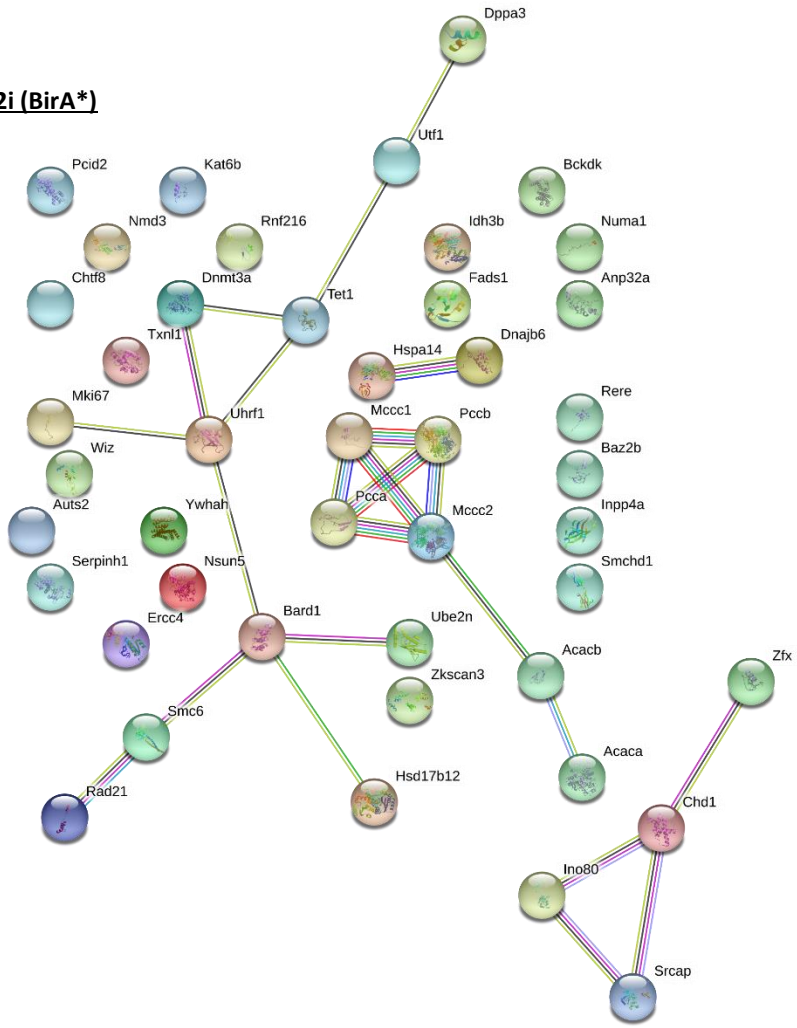
N-terminus Uhrf1 Serum (BirA*)



Biological Process (GO)		
Pathway Description	**Count in gene set**	**False discovery rate**
Organic substance metabolic process	32	5.99×10^{-5}
Primary metabolic process	32	5.99×10^{-5}
Cellular metabolic process	32	5.99×10^{-5}
Nitrogen compound metabolic process	31	5.99×10^{-5}
Nucleobase-containing compound metabolic process	22	6.83×10^{-5}

F

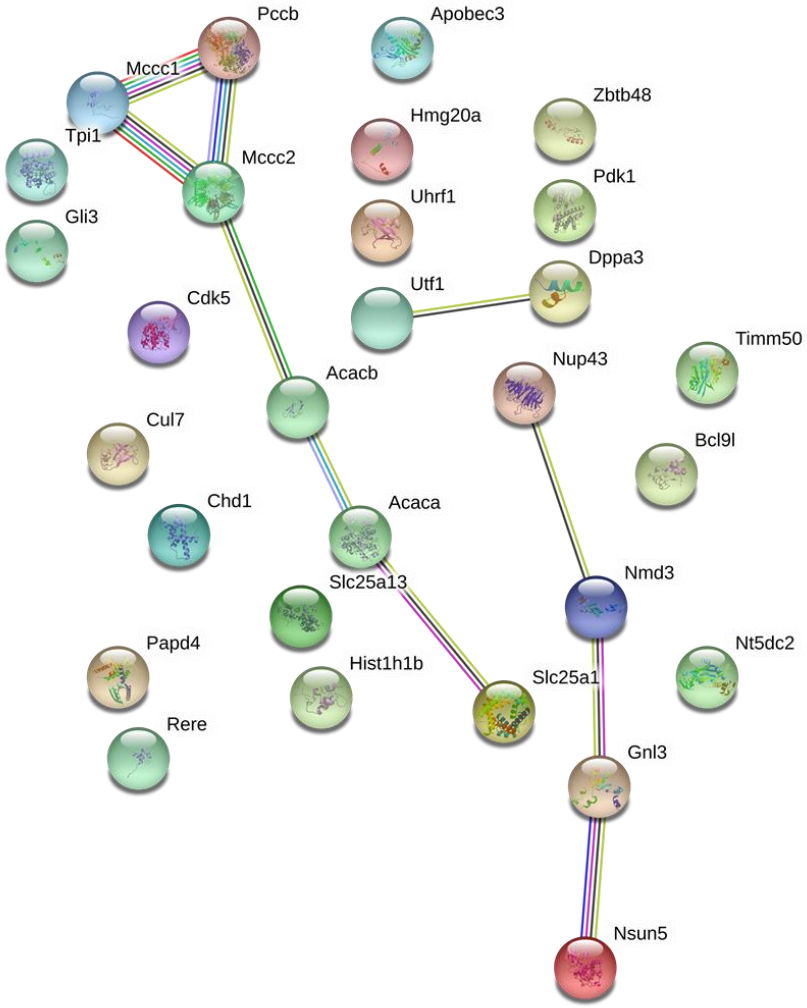
N-terminus
Uhrf1 KOSR + 2i (BirA*)



Biological Process (GO)		
Pathway Description	**Count in gene set**	**False discovery rate**
Chromosome organisation	15	1.61×10^{-7}
Organic cyclic compound metabolic process	27	4.64×10^{-7}
Regulation of chromosome organisation	10	4.64×10^{-7}
DNA metabolic process	12	1.27×10^{-6}
Nucleobase-containing compound metabolic process	25	1.27×10^{-6}

G

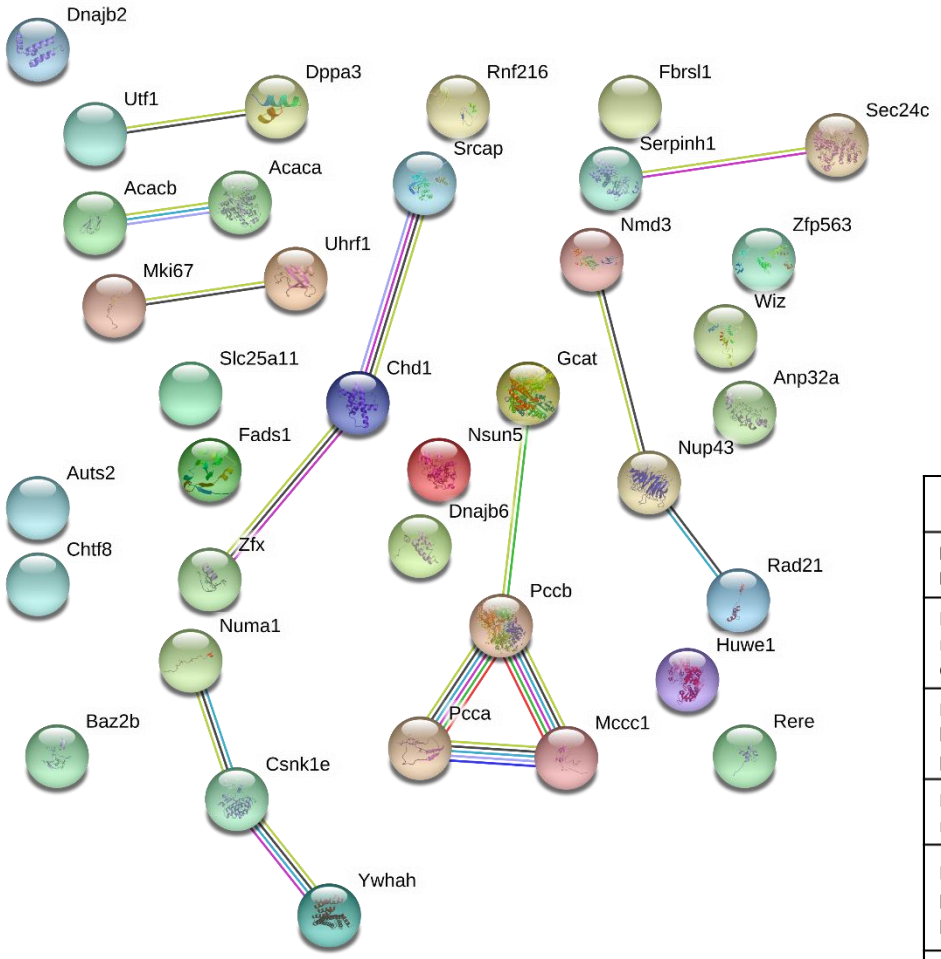
C-terminus
Uhrf1 Serum (BirA*)



Biological Process (GO)		
Pathway Description	**Count in gene set**	**False discovery rate**
Malonyl-CoA biosynthetic process	2	3.90×10^{-3}
Malonyl-CoA metabolic process	2	3.90×10^{-3}
Positive regulation of cellular protein localisation	5	3.90×10^{-3}
Cellular component organisation or biogenesis	16	3.90×10^{-3}
Organic substance metabolic process	22	3.90×10^{-3}

H

C-terminus
Uhrf1 KOSR + 2i (BirA*)



Biological Process (GO)		
Pathway Description	**Count in gene set**	**False discovery rate**
Regulation of mitotic nuclear division	5	5.90×10^{-3}
Malonyl-CoA biosynthetic process	2	6.30×10^{-3}
Malonyl-CoA metabolic process	2	6.30×10^{-3}
Regulation of primary metabolic process	20	6.30×10^{-3}
Regulation of macromolecule metabolic process	20	6.30×10^{-3}

Figure on previous page **Figure 4.7 –** Visual representation of BioID data generated by STRING for each of the noted lines, conditions and enrichment over Empty vector or free BirA* control lines. Inset tables are Gene Ontology terms returned for the proteins displayed. Lines are indicative of a report from different methods that the proteins labelled above each sphere interact or have the potential to.

In both serum and KOSR+2i, proteins enriched over the Empty vector control are associated largely with nuclear processes. These are feasible categories for attempts to identify novel interactors that influence the activity of Uhrf1 (Figure 4.7 A-D). Furthermore, the majority of proteins have pre-established links and those that lay outside the web of interactions, such as the zinc-finger proteins (Zfp), would likely be within close proximity to Uhrf1 as they can bind DNA.

Overall, direct interactors with Uhrf1 are well represented, specifically by Dnmt1, DNA Lig1 and TopII-α, which appears to be a particularly rich node of interactions (Sharif *et al.*, 2007; Lu *et al.*, 2015; Ferry *et al.*, 2017) . In Figure 4.7 A for example, TopII-α has been inferred to interact with the RNA helicase Dhx9 as there are putative homologues that demonstrably associate in other species. Similarly, homologues of Dhx9 interact with other RNA helicases and ribonucleotide proteins, as shown in the vicinity of the Dhx9 node in Figure 4.7 A. As BioID biotinylates vicinal proteins, the fact this interacting network has been identified in the screen may be a reflection that they share a co-factor in TopII-α and remain associated with it, or are in the process of dissociating, when TopII-α facilitates Uhrf1 activity. Histones have also been detected in this dataset and in other media and with use of the C-terminal tag. The histones are illustrated to interact with numerous other proteins, particularly those annotated as helicases or that modify histone tails, such as Artx, Trim28, Smarca5 and Jmjd1c. As Uhrf1 engages with histones during DNA methylation maintenance, as covered in Chapter 1, this may account for a direct or indirect interaction with these examples.

As is the case from the serum data, Rif1 is not annotated to interact with Uhrf1 in the network displayed in Figure 4.7 B, which shows the fifty most abundant proteins in the N-terminus KOSR+2i dataset. Whilst it is shown to likely interact indirectly through Wdhd1, a transcription initiation factor, there has been a report that Uhrf1 and Rif1 cooperate in the DNA damage response (H. Zhang *et al.*, 2016). This demonstrates that STRING can facilitate the processing of data but should not be relied on solely to interpret it.

In the C-terminus data in Figure 4.7 C, the presence of DNA Lig1 and TopII-α in the top fifty most abundant proteins can account for a number of interactions, similar to the N-terminus data. The co-expression of Uhrf1 and Bptf, along with the fact homologues in other species reportedly interact, has led to an annotation between these two proteins. This transcription factor resides within the centre of various different interactions between multiple other proteins, some of which, such as TopII-α and Artx are known or inferred to interact with Uhrf1. Despite the annotated role of Brd4 as a reader of acetylated histones, this protein has not been associated to interact with Uhrf1 directly, but potentially through Bptf. In both cases, the likelihood of Uhrf1 being in proximity to Bptf and Brd4 is high as their functions involve chromatin, the substrate for Uhrf1 binding. Furthermore, as Brd4 also mediates the DNA damage response (X. Li *et al.*, 2018), this context may provide a route in which Uhrf1 and Brd4 interact directly, in light of the known function of Uhrf1 in this process as referred to earlier with Rif1.

As for the N-terminus data, the top fifty most abundant proteins in the C-terminus are broadly similar between serum and KOSR+2i media. However, in Figures 4.7 E-H, which relate to BirA* enriched proteins, the GO terms diverge somewhat between the culture conditions, reflecting a change in the protein complexes that Uhrf1-BirA* may interact with or encounter in the vicinity.

The N-terminus serum media BirA* enriched dataset has been presented as having few associations (see Figure 4.7 E). Those drawn between Uhrf1 are with a cell-cycle marker, Ki-67, which coats chromosomes during mitosis, and a DNA damage mediator, Bard1. This demonstrates the opportunity to identify novel interactors. Compared to the N-terminus, the number of proteins enriched over the free BirA* control in serum media with the C-terminally tagged Uhrf1 is less and all but eight proteins are found in the N-terminus dataset also, hence the similarity between the STRING plots (see Figure 4.7 G).

The larger number of proteins enriched over free BirA* in KOSR+2i media has led to a more extensive web of inferred interactions emanating from Uhrf1. In the N-terminus data, shown in Figure 4.7 F, Tet1 has been inferred to interact with Uhrf1. This may be expected to be detected in the BioID data, given the minor role of active mechanisms, mediated by Tet1 for example, in the demethylation promoted in 2i media (von Meyenn *et al.*, 2016). Furthermore, the SRA domain of Uhrf1 that is typically associated with binding to 5mC, has also been shown to have an appreciable affinity for 5hmC, the product of Tet1 activity (Frauer

et al., 2011). Perversely, Dnmt3a is also enriched over the BirA* control in KOSR+2i, despite the protein declining dramatically in this media, as shown in Chapter 3. However, when accounting for the fold-difference in enrichment between serum and KOSR+2i media, this does not pass the threshold. Notwithstanding this, there is evidence that Uhrf1 can influence *de novo* methylation in oocytes (Maenohara *et al.*, 2017).

The C-terminal data for proteins enriched over the BirA* control in KOSR+2i is largely similar to the N-terminal equivalent. Only eight unique proteins were identified using this line and none of these may reportedly interact with Uhrf1. Superficially, some examples of these such as Zfp563 and Huwe1 could be rationalised to be vicinal to Uhrf1 in mESCs, based on their annotations, as they have the capacity to bind DNA and are involved in the DNA damage response, respectively.

Of note are Pcca/b and Mccc1/2, which are present in all datasets. These are mitochondrial proteins and in the case of Pcca and Mccc1, they have annotated biotin carboxylase domains that transfer CO_2 groups from bicarbonate to biotin in an intermediate step of macromolecule metabolism. The malonyl-CoA processes in the C-terminal BirA* enriched data can also be accounted for by this action as Acaca/b possess a biotin carboxylase domain too (Tong, 2013).

The GO terms of proteins in the Uhrf1 dataset for the serum media Myc pull-down relate to metabolic processes associated with nucleic acids, nucleobase-containing compounds, RNA and cellular nitrogen compounds, in addition to gene expression. For KOSR+2i, two defined categories, regulation of gene expression and nucleobase-containing compound metabolic processes, are returned.

Collectively, this presents a strong case that the Uhrf1-BirA* fusion is localising appropriately and therefore that novel interacting proteins are viable candidates for further functional characterisation.

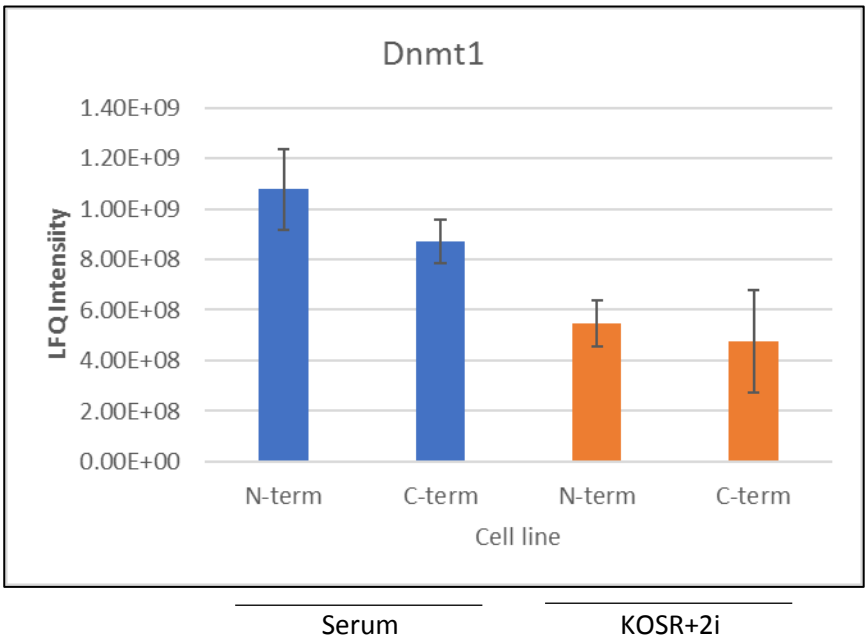


Figure 4. 8 – LFQ intensity of Dnmt1 determined by mass spectrometry in the Uhrf1-BirA* lines in serum and KOSR+2i medias. Error bars represent the standard deviation from the mean of three technical replicates.

The categories for the DNMT1 BioID dataset enriched over the Empty vector control include metabolism of nucleic acids, nucleobase-containing and nitrogen compounds. The small number of BirA* enriched proteins and those from the pull-down precludes categorisation however, 'small molecule metabolic processes' has a low false discovery rate for serum BioID data. Proteins enriched in the DNMT1 Myc pull-down from serum media are examples of chromosome and nucleosome organisation regulators. This may be in part due to the number of histone proteins and histone modifying enzymes in the dataset. Known DNMT1 interactors Pcna and Usp7, which have been addressed in the introduction, are also found. Proteins in the KOSR+2i Myc pull-down dataset are represented by the categories: nucleic acid and RNA metabolic processes and regulation of gene expression. Relevant to this former category is TSG101 which can interact with DNMT1 through DMAP1 and repress transcription, as discussed in the introduction (Rountree, Bachman and Baylin, 2000).

Of note, Dnmt1 was calculated in the Uhrf1 BioID screen to be enriched over the Empty vector control only. Whilst intensity was reduced in the KOSR+2i condition, which is expected given evidence that DNA methylation maintenance is impaired, the relative difference with

serum culture was not sufficient for it to be classified as a differential interacting protein. A 1.97 and 1.82-fold difference was exhibited for the N and C-terminus line respectively (Figure 4.8). Nevertheless, 29 proteins enriched over the Empty vector control in the Uhrf1 BioID data were more abundant in serum than in KOSR+2i, and similarly 185 passed the threshold in KOSR+2i. Of proteins enriched over the BirA* control, this was reduced to 10 in serum compared to KOSR+2i and 9 in the opposite manner.

4.2.3 Selecting candidates for functional characterisation

After surveying the data, five candidates were selected as they showed a different propensity to interact with Uhrf1 in the culture conditions, or for scientific curiosity. In all cases, they were enriched over the most stringent control in the BioID experiments, free BirA*. Furthermore, all were potential novel interacting partners, with the exception of Dppa3/Stella (von Meyenn *et al.*, 2016; Mulholland *et al.*, 2018; Du *et al.*, 2019).

Figure 4.9 and Table 4.2 outline the average LFQ intensities of the proteins across the experimental lines in serum and KOSR+2i culture and the fold-differences. Apobec3 (Apolipoprotein B Editing Complex 3) and Utf1 (Undifferentiated embryonic cell Transcription Factor 1) were notable for their relative enrichments in serum media, in contrast to KOSR+2i. Conversely, Stella interacted more strongly in KOSR+2i than serum. The fold differences between serum and KOSR+2i are presented in Table 4.2 and notably exceed the set threshold. Auts2 also displayed variable levels however, a slightly less stringent *p* value was required to be set for the interaction in serum to be accounted for. The N-terminal value was otherwise filtered out for having a *p* value of 0.053 and the C-terminal line at 0.056.

A number of differential interactors were also found that perform PTMs on other proteins (Figure 4.10). Pdk1 (Pyruvate Dehydrogenase Kinase 1) was selected for characterisation as this displayed the greatest fold difference in enrichment between serum and KOSR+2i media Indeed, Pdk1 was the only example of this class enriched over the free BirA* control and it was found exclusively in the serum BioID data. Furthermore, it was also 15.29-fold more enriched in the DNMT1 C-terminus BioID data and cleared the threshold set for enrichment over the free BirA* control.

The Uhrf1 Myc pull-down data lends some support to the BioID findings. Whilst Stella was not differentially enriched by the criteria imposed, there was concordance with the trend observed with BioID. Similarly, Apobec3 and Pdk1 were more enriched in serum, supporting

the BioID result, but not over the background values obtained with pull-down of the Myc control. Meanwhile, Utf1 was detected in the pull-down from the serum-maintained line only and Auts2 was not found in any sample, including the Myc control line (Figure 4.9 and Table 4.2).

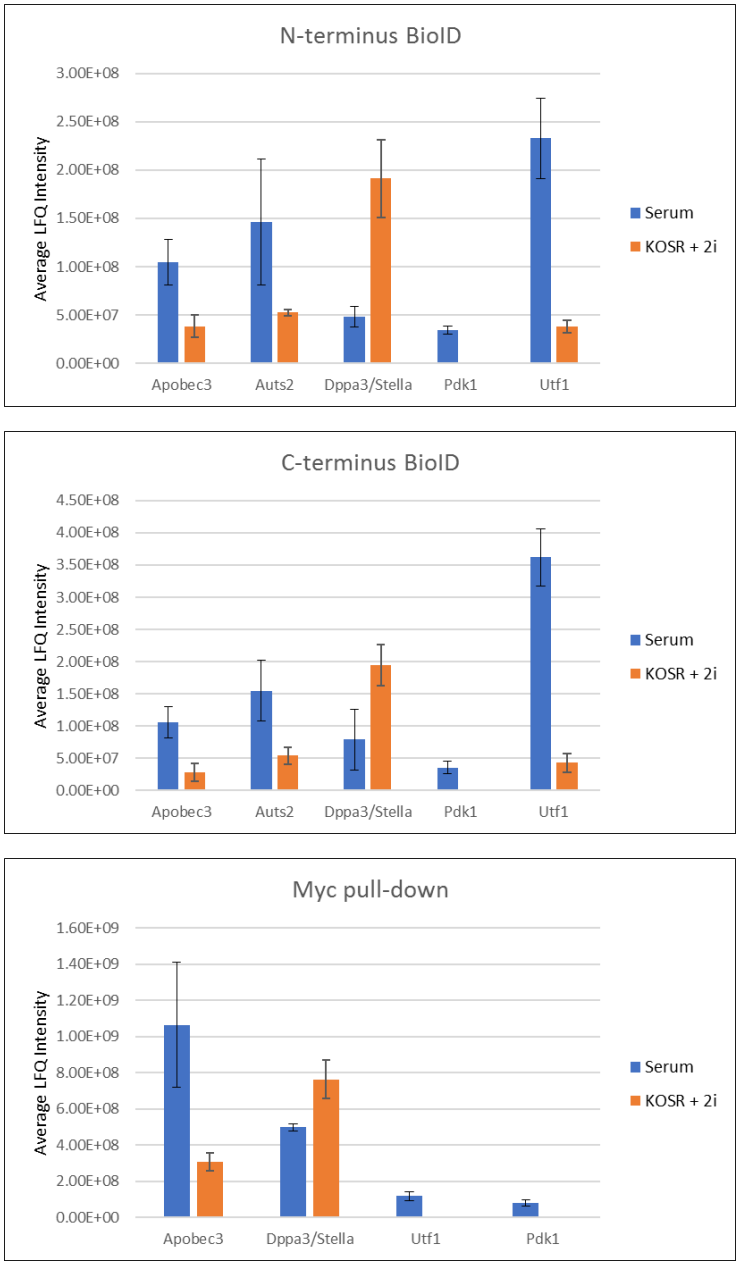


Figure 4.9 – Average LFQ intensity of the candidate proteins in **A**. N-terminally BirA* tagged Uhrf1, **B.** C-terminally BirA* tagged Uhrf1 and **C.** Myc mediated pull down of the N-terminally BirA* tagged Uhrf1 line. Error bars represent the standard deviation from the mean of three technical replicates.

Table 4.2 – Candidates and relative extent of enrichment in BioID and Myc-pull down screens from Uhrf1-BirA* lines cultured in serum and KOSR+2i media

Protein	BioID	Differential enrichment	Myc pull-down
Apobec3	N & C –term	X	Not enriched over background
Auts2	N & C –term	Under threshold	X
Dppa3/ Stella	N & C –term	More enriched in KOSR + 2i N-term: 3.96-fold C-term: 2.47-fold	More enriched in KOSR + 2i 1.53-fold
Pdk1	N & C –term (not in KOSR+2i)	More enriched in Serum N-term: 3.41 x 10^7-fold C-term: 3.54 x 10^7-fold	Not enriched over background
Utf1	N & C –term	More enriched in Serum N-term: 6.10-fold C-term: 8.43-fold	More enriched in Serum (not detected in KOSR+2i) 1.19 x 10^8-fold

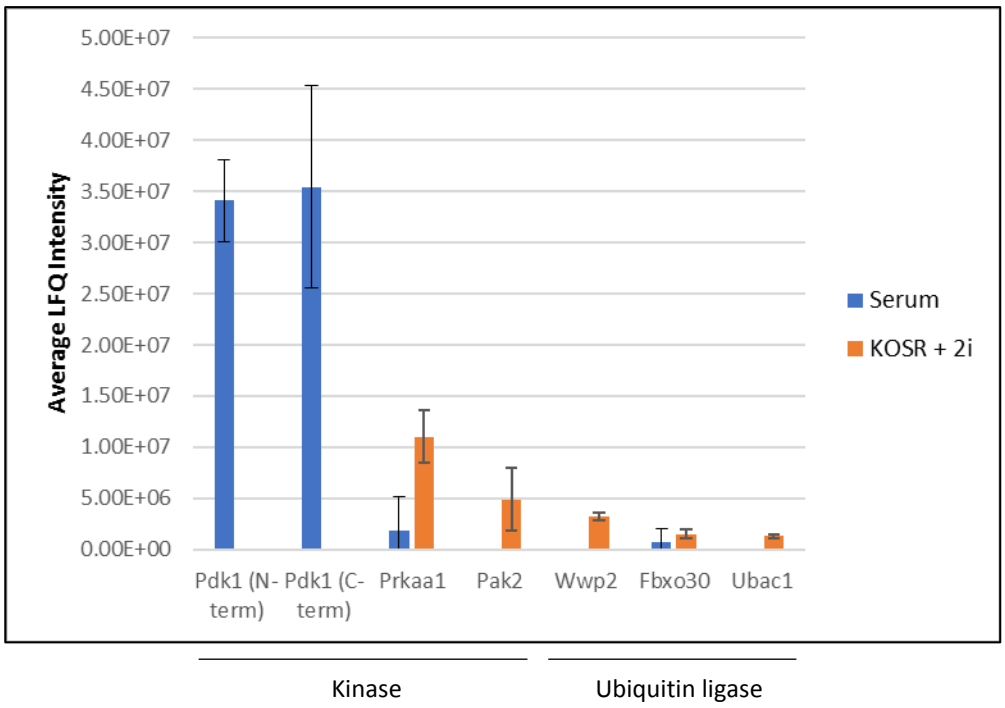


Figure 4.10 – Average LFQ Intensities of differential interacting proteins that mediate post-translational modifications. Candidates were found exclusively in the N-terminus data with the exception of Pdk1 which was detected at the N-and C-terminus. Pdk1 was the only example enriched over the free BirA* control. Error bars represent the standard deviation from the mean of three technical replicates.

To investigate whether the variable interactions could be explained by differing levels of the candidate proteins in the media, I performed a western blot and RT-qPCR for each (Figures 4.11 and 4.12). I was however unable to validate an antibody for Stella. For all candidates, the changes in expression and protein levels correlate with the extent of interaction in the different media. Both Utf1 and Pdk1 are subject to downregulation in KOSR+2i and this is accompanied by a reduced interaction with Uhrf1. Meanwhile Stella increases in expression at the mRNA level in KOSR+2i, matching the more extensive interaction with Uhrf1 that was observed. Whilst Apobec3 did not reach the threshold to be defined as a differentially interacting protein, the western blot shows levels decline subtly in KOSR+2i and there was a trend to reduced binding in the same condition. Lastly, Auts2 levels remain consistent across the media and a differential interaction cannot be confirmed as the LFQ intensity fold difference (2.78 and 2.88 for N- and C-term lines respectively) falls below the imposed threshold.

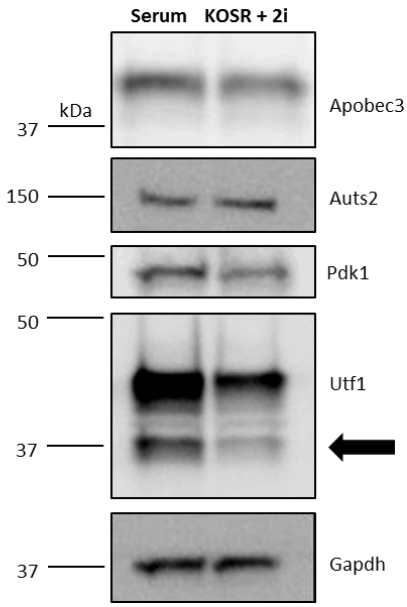


Figure 4. 11 – Western blot of candidates in mESCs maintained in serum-based media and after 48 hours in KOSR+2i. The upper band in the Utf1 blot is of a non-specific band reported by the manufacturer. Gapdh has been used as a loading control

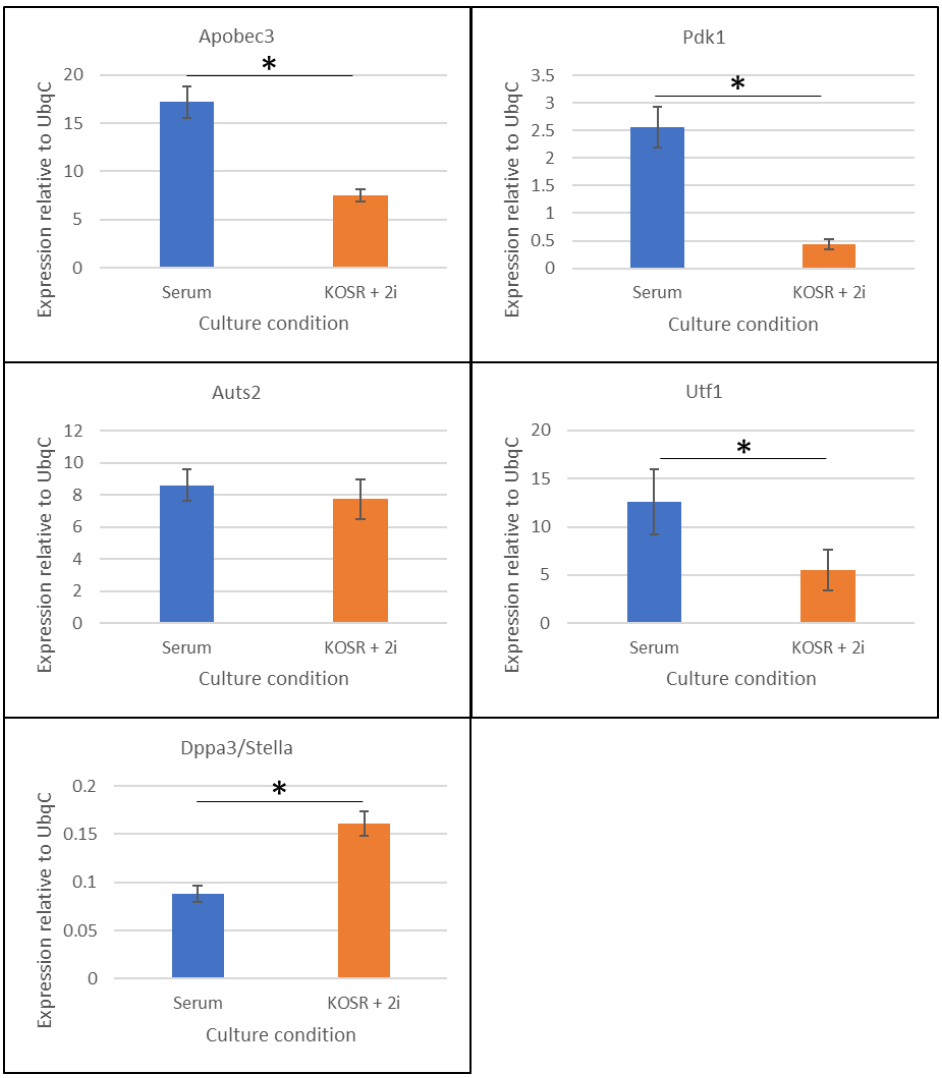


Figure 4.12 – Expression of the selected candidates in serum and KOSR+ 2i media relative to *UbqC*, as determined by RT-qPCR. Error bars represent the standard deviation from the mean of three technical replicates. (* = $p<0.05$)

In an initial attempt to validate the mass spectrometry data, I repeated the BioID and Myc pull-down experiments and probed for the proteins of interest in a western blot (Figures 4.13 and 4.14). In addition, I performed an IP of endogenous Uhrf1 in a WT E14 line (Figure 4.15). In all cases, there was no concordance with the dataset as the candidates were not enriched in the experimental samples. Whilst interactions may not be preserved by the Myc pull-down and IP due to their transient nature, I was unable to recapitulate the BioID data too. One possible explanation is that the proportion of Uhrf1 that interacts is low to the extent that it

is below the detection limit of a western blot, meaning alternative methods to validate the candidates need to be explored as discussed below.

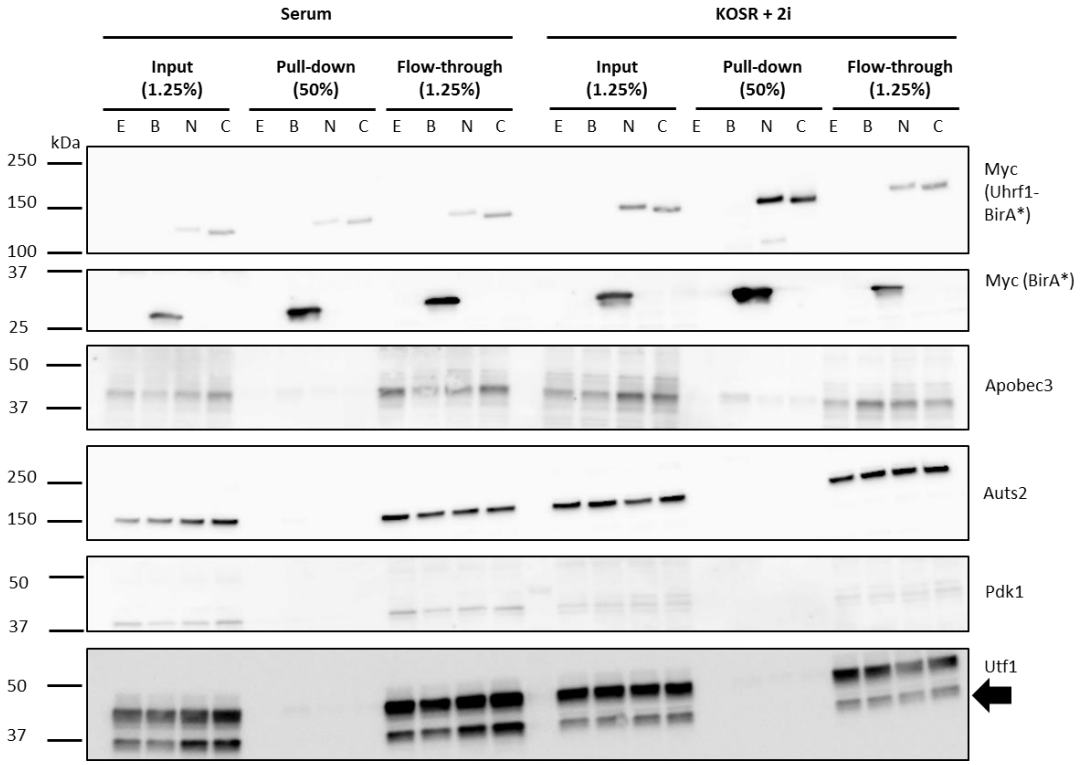


Figure 4.13 – Western blot of candidate interactors following treatment of the Uhrf1-BirA* lines and controls with biotin in the respective medias and streptavidin mediated pull-down. Empty vector control (E), Free BirA* control (B), BirA* N-terminally tagged Uhrf1 (N) , BirA* C-terminally tagged Uhrf1 (C).

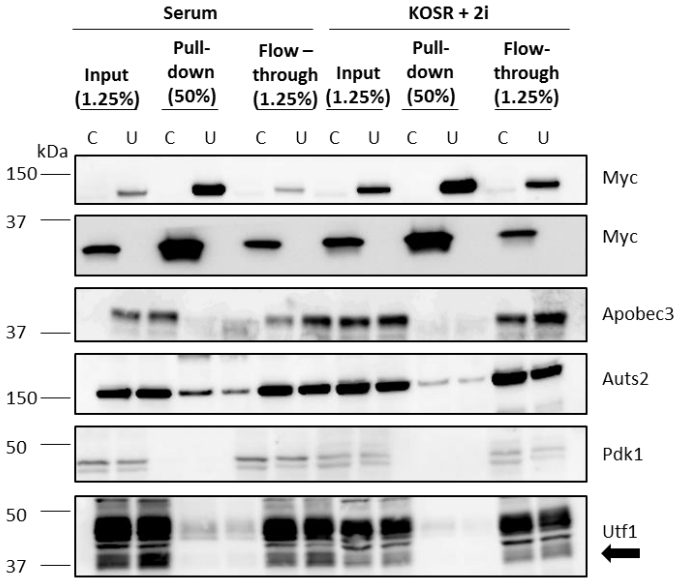


Figure 4.14 – Western blot of the candidate interactors following Myc-tag mediated pull-down of N-terminally BirA* tagged Uhrf1 and control lines cultured in serum or KOSR+2i media. Myc pull-down control (C), N-terminally BirA* tagged Uhrf1 (U)

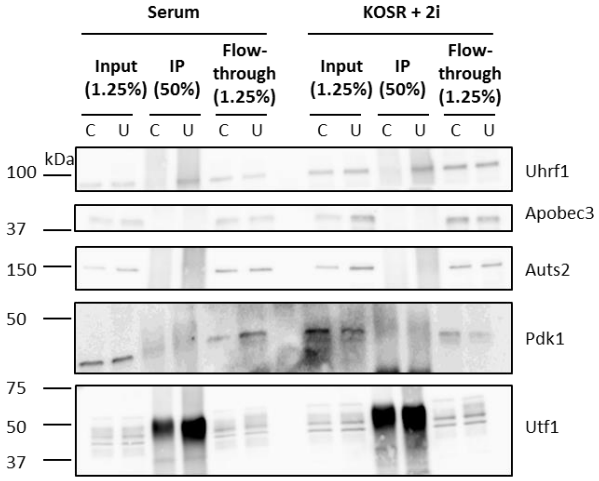


Figure 4.15 – Western blot of candidate interactors after endogenous Uhrf1 immunoprecipitation from a WT E14 line maintained in serum or KOSR+2i media. IgG control (C), Uhrf1 antibody (U).

4.3 Discussion

Using a proximity labelling method and a high affinity pull-down, I have generated datasets that show the differential interactome of Uhrf1 in medias that maintain a high level of DNA methylation or promote a global loss of the mark. As Uhrf1 appears central to this epigenetic remodelling, the interacting proteins in these markedly different signalling environments may illustrate mechanisms that contribute to this process. Similar attempts to investigate the profile of DNMT1 have proven to be less successful and therefore a repeat would be necessary to obtain data of comparable quality. Led by the proteomic data, five candidates were chosen for further characterisation and to determine whether an interaction with Uhrf1 is instructive in regulating DNA methylation.

To this end, initial experiments would investigate the kinetics of DNA demethylation after genetic perturbation of each candidate. Proteins that interact more strongly in serum may be expected to delay demethylation when overexpressed, with an opposite effect when the candidate is knocked down and the cell line transitioned to KOSR+2i, and vice versa. Whether knockdown of any of the factors influences demethylation in KOSR+2i in conjunction with Uhrf1 perturbation should also be followed. As knockout of *Uhrf1* will immediately begin to promote DNA hypomethylation (Sharif *et al.*, 2007), a conditional line would prove to have more utility for temporal control. Alternatively, a combined knockdown may be sufficient.

The higher levels of the DNMT1- and Uhrf1-BirA* fusions in KOSR+2i suggest that the media promotes a faster rate of proliferation, compared to serum culture (Figures 4.3 and 4.4). Although this has not been tested thoroughly, I noticed this when passaging cells maintained in KOSR+2i. In order to identify differential interactors, the data from the screens required further normalisation and an adjusted threshold limit. To minimise the manipulation of data and improve future screens, the protein concentration in the cell lysates could be quantified after harvesting the treated cells and then equal amounts from each of the serum and KOSR+2i samples taken to perform the respective pull-downs.

I have made preliminary steps to validate the interactions between Uhrf1 and the candidates selected for functional characterisation however, so far this has proven to unsuccessful. In a study focussed on core histones, affinity-purification and BioID experiments were performed in parallel and both approaches yielded diverse profiles of interacting proteins. BioID was notable for detecting lowly abundant proteins, which may partly explain the results of the

validation attempts for my screens (Lambert *et al.*, 2015). An alternative and frequently used method to confirm novel interacting partners from BioID datasets is immunofluorescence for visualising co-localisation with the protein of interest (Varnaitė and MacNeill, 2016). Whilst this does not strictly determine if a protein directly interacts, this can also be argued for BioID as this process surveys the vicinity of the tagged protein from which inferences are then made. As set out in Table 1.1, overexpression of the protein in an unrelated cell line has proven an appropriate method for confirming a functional interaction between Uhrf1 and Dnmt1. This approach can be modified by performing sequential co-IPs with truncated proteins in order to identify the sites that are important for mediating the interaction.

Nevertheless, the fulfilment of the five chosen candidates to the strictest control in the BioID experiment provides an impetus to investigate them further. The biological roles of the candidates are discussed below and their potential relationships with DNA methylation maintenance are hypothesised.

4.3.1. Dppa3/Stella

The immediate stages of early development that occur after fertilisation are characterised by pervasive DNA demethylation. This affects the maternal and paternal genomes asymmetrically, as the maternal counterpart demethylates relatively more slowly (Smallwood and Kelsey, 2012). One factor that has been suggested to mitigate this effect is Stella, also known as Dppa3, which prevents active demethylation mechanisms from remodelling the maternal genome (Nakamura *et al.*, 2012). It is therefore surprising that Stella has also been shown to disrupt the activity of Dnmt1 and promote demethylation in oocytes and mESCs (Mulholland *et al.*, 2018; Y. Li *et al.*, 2018). In the Uhrf1 interactome screens, Stella proved to be one of the most robust candidates as it was enriched over the free BirA* control and was also detected in the Myc pull-down data. Furthermore, in keeping with the notion that Stella is associated with hypomethylation, it was relatively more enriched in the KOSR+2i condition.

It has been shown in oogenesis that Stella is evenly distributed between the nucleus and cytoplasm in growing oocytes and is more enriched in the former compartment at later stages. Meanwhile, Uhrf1 is mainly maintained in the cytoplasm. In a *Stella* knockout model however, Uhrf1 ectopically accumulates in the nucleus. Subsequently, there is a corresponding increase in DNA methylation across a range of genes and repetitive sequences, which is most notable in mature (MII) oocytes. As a consequence, the maternal

genome is hypermethylated at pronuclei at stages immediately post-fertilisation and the two-cell stage, wherein transcriptional defects begin to emerge. Using separate *Dnmt3a* and *Dnmt1* null MII oocytes in a *Stella* null background, over 80% of the aberrantly methylated regions that arise in *Stella* deficient oocytes were shown to be attributable to Dnmt1 activity. The influence of Dnmt3b was not tested as it was undetectable at this stage. This study not only broadens the role of Stella in pre-implantation development, it also suggests that Dnmt1 has a *de novo* methylation function. The typical Dnmt1 substrate is regarded as newly synthesised hemi-methylated DNA however, it was shown with *Stella*-null oocytes that aberrant patterns arose when they were mitotically arrested (Y. Li *et al.*, 2018).

Further evidence of a relationship between Stella and Uhrf1 can be found in a study that initially began dissecting the catalytic and catalytically independent roles of Tet1 in mESCs (Mulholland *et al.*, 2018). Gene expression analyses in catalytically dead Tet1 and *Tet1*KO mESC lines and a WT EpiLC line, which are all markedly hypermethylated compared to WT ESCs, showed that *Stella* was notably lower compared to WT ESCs. The propensity for Tet1 to bind to a transcriptional enhancer of *Stella* and an enrichment of 5hmC at this element was later shown. Collectively this suggests that *Stella* transcription is influenced by Tet1 activity. One of the main functions of Tet1 in pre-implantation development was therefore hypothesised to be to actively demethylate regulatory elements of *Stella* and promote its expression. In turn, deletion of *Stella* in mESCs was shown to increase 5mC levels by ~25%. Analogous to the study above, ectopic expression of *Stella* was observed to promote the translocation of a Uhrf1-GFP fusion protein out of the nucleus and into the cytoplasm in a WT line maintained in N2B27+2i. Similarly, a Dnmt1-GFP fusion had disrupted chromatin targeting and a small fraction was localised to the cytoplasm (Mulholland *et al.*, 2018). Notably, this challenges other reports that have used the same conditions but did not observe the redistribution of Uhrf1 (von Meyenn *et al.*, 2016).

The evidence presented above strongly suggested that the differential interaction observed in the screens I performed is functional, in that Stella may drive Uhrf1 out of the nucleus and as a consequence impair Dnmt1 targeting. I therefore tested this by visualising the distribution of Uhrf1 in a WT E14 line maintained in serum or KOSR+2i for 48 hours.

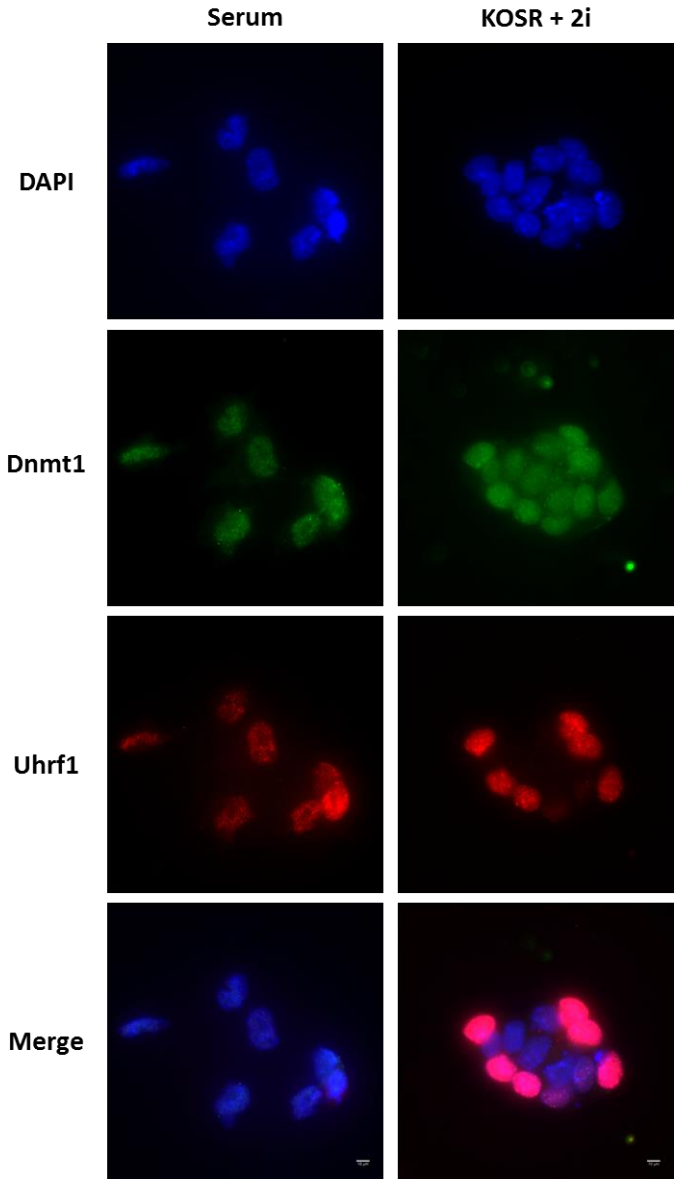


Figure 4.16 – Immunofluorescence images of Dnmt1 (green channel) and Uhrf1 (red channel) (and DAPI (blue channel)) in mESCs cultured in serum-based or KOSR+2i media for 48 hours. A total of six images of different colonies were taken of each line. Scale bars represent 10µm.

Figure 4.16 shows that the result is inconclusive due to the propensity of the antibodies to give a strong background signal however, the localisation of Uhf1 and Dnmt1 does appear to be comparable between the different media. Whilst Stella may indeed promote the translocation of Uhrf1 at later stages, thus compounding the effect of low Uhrf1 protein abundance, other mechanisms may initiate the impairment of DNA methylation maintenance. However, aside from an ability to sequester Uhrf1, Stella can reportedly influence DNA demethylation independently of this function. From a series of *in vitro* peptide binding assays, the C-terminal end of Stella was shown to compete with histone tails in binding the TTD and PHDs of Uhrf1 (Du *et al.*, 2019). A similar effect could be tested in KOSR+2i by fractionating WT and *Stella* knockout mESCs and assaying for the amount of Uhrf1 bound to chromatin or free in the nucleoplasm. Experiments from the aforementioned studies could also be adopted, including the use of 5-aza-CdR to trap Dnmt1 and quantify the amount bound to chromatin in the two culture systems, or Fluorescence Recovery After Photobleaching (FRAP) to ascertain whether Dnmt1 and Uhrf1 localisation occurs with similar kinetics (Mulholland *et al.*, 2018; Du *et al.*, 2019).

4.3.2 Utf1

A candidate that has an inverse profile to Stella in the manner in which it interacts with Uhrf1 is Utf1. This was initially characterised as transcriptional regulator of high importance in development and is expressed in PGCs, the ICM and ESCs. It is transcriptionally regulated by Oct4 and Sox2 and as such, this may explain its downregulation in post-implantation development (Okuda *et al.*, 1998; van den Boom *et al.*, 2007; Bao *et al.*, 2017). Furthermore, knockdown of this factor negatively affects ESC differentiation as exemplified by the maintenance of Oct4 and Nanog levels and misexpression of lineage markers when forming embryoid bodies. Native ChIP identified genes to which it binds and these are involved in chromatin, nucleosome and protein-DNA complex assembly. In addition, it was enriched on the Y-chromosome at genes associated with spermatogenesis. Collectively, only ~6% of these bound targets were differentially expressed upon *Utf1* knockout (Kooistra *et al.*, 2010). Another study in ESCs, where biotin-ChIP-seq with a biotin-Utf1 protein was performed, noted that the density of CGIs was high at Utf1-bound TSSs, relative to the genome average (Jia *et al.*, 2012).

Utf1 was shown to have an affinity for chromatin comparable to core histones, leading to the speculation that it may assist in establishing a permissive epigenetic state to enable

differentiation (van den Boom *et al.*, 2007). As set out in the introduction however, DNA methylation of genes may only be acquired after transcriptional silencing to lock in this regulation. Whether Utf1 therefore sets up a favourable epigenetic state in response to or before transcription needs to be temporally dissected (van den Boom *et al.*, 2007).

Nevertheless, a role in modulating epigenetic marks has been ascribed to Utf1 (Jia *et al.*, 2012). Utf1 peaks were shown to correlate highly with those of a PRC2 component (Suz12) in ChIP-seq experiments in mESCs. Furthermore, PRC2 binding increased up to 4-fold along with H3K27me3 deposition upon *Utf1* knockout. This led to the conclusion that Utf1 acts to regulate the polycomb system by restraining excessive activity. Rescue experiments and *in vitro* assays with PRC2 components and Utf1 demonstrated there is competition for a similar DNA motif enriched at bivalent domains, genomic regions at which both Utf1 and PRC2 are enriched. Further experiments also showed that Utf1 acts to assist in the efficient loading of mRNA decapping components onto nascent mRNAs at these regions, thus pruning the expression of bivalently marked genes too. As such, genes that were upregulated upon Utf1 loss were examples of those escaping pruning whereas downregulated genes, marked by higher CpG densities, accumulated PRC2 and were silenced (Jia *et al.*, 2012) .

Uhrf1 was also shown to have a regulatory role at bivalent domains and therefore the interaction detected in the BioID and Myc pull-down screens could reflect a co-occupancy with Utf1 (Kim *et al.*, 2018). In a *Uhrf1*-null mESC line, over half of the bivalent domains that were detected in a WT line resolved to either an active or repressed state. Mechanistically, the deposition of H3K4me3 borne by active elements and bivalent domains was attributed to the direct recruitment of Setd1a by Uhrf1, an interaction I also found to occur in the C-terminally tagged Uhrf1 in serum media. Furthermore, Uhrf1 was shown to modulate the binding of Oct4 at regions marked with H3K4me3. This may account for the observation that when *Uhrf1* is transduced with *OSK* into MEFs, the efficiency of reprogramming cells to an induced pluripotent state is increased 3-fold, when compared to the use of *OSK* alone. The number of reprogrammed colonies produced by *OSK*+*Uhrf1* is notably comparable to that attained using the standard *OSKM* cocktail (Kim *et al.*, 2018). Of note, Klf4 was found to bind to Uhrf1 in my dataset too. This would be an interesting avenue to explore regarding the establishment of pluripotency by Uhrf1.

Aside from histone marks, an association between Utf1 and DNA methylation has also been observed. Whether this effect is direct or indirect needs to be determined. When knocked

out, H3K9me3 levels, which are highly associated with Uhrf1, remained consistent yet demethylation of minor satellite regions was shown (Kooistra *et al.*, 2010). Evidence from the BioID and Myc pull-down screens showing Utf1 binds to Uhrf1 suggest that the effect of *Utf1* knockout on DNA methylation may be direct.

Complicating future experiments to determine a functional relationship with Uhrf1 is the fact that knockdown of *Utf1* causes cells to proliferate more slowly compared to a WT line. This is due to an increase in levels of the tumour suppressor *Arf*, which is a likely consequence of impaired mRNA pruning that Utf1 would otherwise mediate (Jia *et al.*, 2012). When maintained in KOSR+2i this may give the impression that demethylation is delayed as DNA methylation will be diluted more slowly when the maintenance machinery is impaired. The effects on the cell-cycle in this medium will therefore need to be characterised before commencing with the experiment. As *Arf* expression is driven by Myc and levels of *Myc* were shown to dramatically decline in N2B27+2i, the effect may be overcome (Ying *et al.*, 2008; Jia *et al.*, 2012). A *Utf1* overexpressed line maintained in KOSR+2i, compared to a WT line in the same medium, might offer an alternative approach to address the same question.

Nevertheless, *Utf1* knockdown has already been shown to promote hypomethylation and as it interacts less extensively in KOSR+2i, evidence suggests it is a robust candidate to follow for further mechanistic characterisation.

4.3.3. Pdk1

The influence of metabolism on DNA methylation may offer a rich seam to mine for the development of novel therapeutics. The generation and recycling of the methyl donor group, Adomet, is after all the product of a metabolic pathway (Ulrey *et al.*, 2005; Kim and Yeom, 2018). In this study, I have found that Pdk1, an enzyme that cooperates with other Pdk isoenzymes to modulate the activity of the pyruvate dehydrogenase complex (Pdc) (Behal *et al.*, 1993), interacts with Uhrf1 exclusively in serum media. Furthermore, it was shown in the DNMT1 C-terminus BioID data to be 15.29-fold more enriched with serum culture compared to KOSR+2i. Bearing a catalytic domain with a similar fold as members of the ATPase/kinase superfamily, Pdk1 modifies serine residues of Pdc (Bowker-Kinley and Popov, 1999).

The decarboxylation of three-carbon compounds by Pdc is a prerequisite for further metabolic downstream processing and the activity of this enzyme therefore needs to be responsive to the requirements of the cell. The role of an inactivating kinase and activating

phosphatase is to integrate environmental cues and act on Pdc accordingly. Longer term regulation however is mediated by Pdh kinase of which Pdk1 is an isoenzyme (Bowker-Kinley and Popov, 1999). A role of Pdc outside of this context has been shown previously in the synthesis of acetyl-CoA in the nuclei of A549 renal- and 786-O alveolar- cancer cell lines. Here the products are utilised by acetyltransferase enzymes to modify histones (Sutendra *et al.*, 2014). A redistribution of Pdk enzymes however was not shown to occur similarly to regulate Pdc. Taking an unbiased approach with regards to the data from the screen would not preclude Pdk1 from consideration as a candidate however.

As introduced above, a network of metabolic pathways influence DNA methylation and therefore perturbations in Pdk1 may indirectly affect the levels of the mark. An approach to circumvent this and define whether the influence is direct would be to map the potential phosphorylation sites on Uhrf1 and Dnmt1 using overexpression models in cells coupled with IPs. This could be followed by a series *in vitro* assays that were similarly used to characterise the modifications set out in Table 1.2. Complementary experiments would be to inhibit the phosphorylation or to generate a phosphomimic before observing the impact on DNA methylation in a cell line. As the enrichment of Pdk1 is in serum media, phosphorylation by this enzyme could be expected to promote DNA methylation, an effect shown to be imparted by other phosphorylations (Table 1.2).

4.3.4 Apobec3

Apobec3 is demonstrably important for restricting the replication of viruses. The protein can be incorporated into viral particles and a DNA editing function is employed to edit cytosine bases to uracil on the minus strand of the viral DNA that is produced by reverse transcription. This subsequently causes G-to-A mutations in the coding strand which frequently leads to missense mutations and the degradation of reverse transcribed DNA (Okeoma *et al.*, 2007; Stavrou and Ross, 2015). In addition, Apobec3 may also act on endogenous retroviral sequences (Esnault *et al.*, 2005). The role of this candidate however may not be restricted solely to this function.

A contribution of active mechanisms in the epigenetic reprogramming of the zygote was assumed due to the faster rate at which the paternal genome is demethylated (Smallwood and Kelsey, 2012). As well as the iterative oxidation of 5mC to an extent that DNA repair mechanisms are signalled to restore a cytosine base, deamination of 5mC to a thymine by Apobec3 has also been implicated to this end. Like oxidation, the aim of this process would

be to invoke DNA repair to restore a cytosine (Nabel *et al.*, 2012). It was therefore surprising that Apobec3 was more abundant in the serum condition, compared to KOSR+2i, in the BioID data. A similar trend was observed in the Myc pull-down but below the value for the background. In both cases however, the differing intensities did not reach the threshold to be regarded as differentially interacting in the medias (Figure 4.9, Table 4.2). Whilst the lower intensity in KOSR+2i could be accounted for by a reduction in Apobec3 levels (Figure 4.11), the remaining protein is still highly abundant and could mediate demethylation regardless.

The function that therefore may appear to unify these roles is the regulation of viral sequences by DNA methylation. Apobec3 may promote expression yet minimise the activity of viruses to an appropriate level to still increase genetic diversity (Esnault *et al.*, 2005). The base media used in the BioID study was composed of 10% KOSR with 1% FCS and the former component with 2i has been shown to induce the expression of genes associated with the 2-cell state, which is partly dependent on endogenous retrovirus activity (Hackett *et al.*, 2017). Analysis of the retroviruses in a subpopulation of mESCs expressing 2-cell stage embryo markers shows them to be hypomethylated and expressed (Macfarlan *et al.*, 2012). In this study, the main driving force for their derepression however was attributed to altered histone modifications. Furthermore, there are differing reports on the extent to which DNA methylation influences the expression of this class, with family-specific effects likely accounting for this and the imposition of stage specific regulatory mechanisms (Karimi *et al.*, 2011; Schoorlemmer *et al.*, 2014; Schlesinger and Goff, 2015).

In any case, the association between a co-factor that is involved in maintaining DNA methylation and another that may remove it seems contradictory. Apobec3 and Uhrf1 interplay therefore presents an interesting focus of study.

4.3.5 Auts2

The Uhrf1 BioID screen identified a number of proteins that are mutated in disease. The nucleosome remodellers Lsh and Cdca7 act in a bipartite complex and influence the methylation of DNA satellite sequences. Mutations that prevent their interaction, and that are present in the *de novo* methyltransferase Dnmt3b itself, are attributed to the development of ICF syndrome, as set out in the introduction (Jenness *et al.*, 2018). Whilst Lsh was detected in both media, a potential interaction with Cdca7 was found exclusively in KOSR+2i, suggesting altered cell-signalling had induced a rearrangement of the complex. That Uhrf1 may interact with Lsh however may be unsurprising in light of data showing Dnmt1

and Lsh can themselves co-IP when overexpressed in a human cell-line (Dunican, Pennings and Meehan, 2015). The relationship between Lsh and DNA methylation and the result of a neurological deficiency when this is disrupted, provided the motivation to select Auts2 for further investigation.

Auts2 was found in the BioID screen to be enriched over the BirA* control in the N and C-terminus of Uhrf1, with protein and mRNA consistent between the two media (Figures 4.9, 4.11 and 4.12, Table 4.2). Although not all of the proteins that reportedly form a stable complex with Auts2 were found in the screen, Rybp was a potential interacting partner of the C-terminus of Uhrf1 in KOSR+2i.

As the name suggests, Auts2 is strongly implicated in autism and mental retardation and the gene is frequently disrupted through translocations in these disorders (Sultana *et al.*, 2002; Kalscheuer *et al.*, 2007). It is highly expressed in the neocortex of mice in early development whilst postnatal levels are lower yet more ubiquitous in the brain (Gao *et al.*, 2014).

Auts2 has been shown to participate in a polycomb PRC1 complex (Gao *et al.*, 2014). Whilst PRC1 complexes are typically associated with transcriptional repression, those featuring Auts2 were first shown to exhibit transcriptionally activating effects in a GAL4-DNA assay. An additional feature of Auts2 containing PRC1 complexes is that they can also operate independently of RING1B, a component that would otherwise ubiquitinate histones to instigate transcriptional repression. A ChIP-seq analysis in mouse brain showed an enrichment of Auts2 at transcriptional start sites, accompanied by PolII and histone marks associated with transcriptional activation. However, the overlap with RING1B-bound genes and Auts2 was described as poor, with ~36% showing a colocalisation of peaks (Gao *et al.*, 2014).

In a *Lsh*-null neural stem/progenitor cell model, compromised self-renewal and lineage commitment along with increased apoptosis have been observed (Han *et al.*, 2017). Mechanistically this has been attributed to impaired nucleosome remodelling, an alteration in histone marks and a reduced accessibility to underlying chromatin. Ultimately this leads to gene expression changes. Aberrant levels of cell cycle regulators and growth factors are likely contenders that account for the downstream effects. As histones and their modifications instruct DNA methylation, loss of *Lsh* would disrupt an early stage in the hierarchy of DNA methylation control. Nevertheless, reduced DNA methylation was found at an enhancer of *Bmp4*, a regulatory growth factor of neural stem cells (Han *et al.*, 2017). A mESC model

expressing a truncated form of *Auts2* meanwhile showed a propensity to apoptose more frequently when promoted to differentiate down a neuronal lineage (Monderer-Rothkoff *et al.*, 2019). The manifestation of this is seen in an *Auts2*-null mouse model, which exhibits developmental delay in a battery of behavioural tests. As nucleosome remodelling activity has not been assigned to Auts2, the compromised neural development is therefore most likely to be accounted for by an influence on epigenetic marks. Indeed, Gene Ontology terms for genes associated with Auts2 ChIP peaks relate to abnormalities in the forebrain and cerebellum (Gao *et al.*, 2014).

The configuration of Auts2 complexes is likely diverse (Gao *et al.*, 2014) and as such, this may impose differing outcomes. The positive influence on gene expression in the GAL4 assay referred to earlier was shown by ChIP to occur when Auts2 was complexed with RING1B (Gao *et al.*, 2014). However as addressed above, evidence suggests that Auts2 can be targeted to chromatin independently of RING1B and therefore the transcriptional effects of the complex when this component is absent may differ. Recently, different Auts2 isoforms have also been identified: a long form is present in mESCs however this is replaced with a shorter version when promoted to differentiate down a neural lineage. Both types are present in adult brain tissue but the short isoform predominates. In humans, a short transcript is expressed in early neurodevelopment and both are detected in the adult brain. Interestingly, assays suggest that the longer isoform associates with transcriptional repression, contrasting sharply with the shorter protein. Furthermore, overexpression of the long isoform in mESCs delays neuronal differentiation, seemingly by promoting *Oct4* and *Nanog* expression (Monderer-Rothkoff *et al.*, 2019). In light of the above, what could be the implications of an interaction with Uhrf1 on DNA methylation?

It is already known that DNA methylation can act instructively with regards to H3K27me3 deposition as hypomethylation promotes a dispersal of the mark across the genome (King *et al.*, 2016). Furthermore, a redistribution of H3K27me3 is evident in 2i media with a N2B27 base (Marks *et al.*, 2012). On the other hand, if Uhrf1 and Auts2 are in a complex with Rybp, the presence of this latter partner is associated with an H3K27me3 independent mode of chromatin targeting and ChIP analyses have not detected an enrichment of H3K27me3 at Auts2 binding sites (Gao *et al.*, 2014). A relationship therefore cannot be drawn directly between H3K27me3 and DNA methylation in this context and will require investigation. An effect of Auts2 in mESCs may also manifest when cells commit to differentiation: WT and

mutant Auts2 lines are initially indistinguishable until promoted to differentiate to cortical neurones (Monderer-Rothkoff *et al.*, 2019). As DNA methylation is influenced by transcription (Neri *et al.*, 2017), a sequence of which will be set off during this process, distinguishing between the direct and indirect effects will be difficult, requiring a specific disruption of a potential Uhrf1 and Auts2 interaction in order to do so.

As a possible complex between Uhrf1 and Auts2 forms in both culture conditions, the rate at which hypermethylation is re-established when WT and *Auts2* overexpressing lines maintained in KOSR+2i are transitioned back to serum-based media could be examined. This will determine whether the components in both media impart an Auts2 dependent effect on DNA methylation. The observation that Auts2 overexpression induces *Nanog* and *Oct4* levels and that an interaction between Auts2 and Uhrf1 is detectable in the KOSR+2i media, a model of ground-state pluripotency, is an intriguing correlation. Returning the cells to serum media also provides an opportunity to investigate whether overexpressed Auts2 is able to counteract the transcriptional heterogeneity associated with *Nanog* in this medium (Marks *et al.*, 2012).

In summary, five candidate regulators of DNA methylation maintenance have been selected based on the proteomic data from Uhrf1 BioID and pull-down experiments. The primary focus is to ascertain whether the interactions influence the changes in DNA methylation that are observed upon transitioning mESCs from serum media to KOSR+2i. If indeed the candidates do, this will illustrate the relationship between a change in protein complex formation and protein activity, in response to differing signalling cues.

Chapter 5 – Investigating the role of citrullination in the regulation of DNA methylation

5.1 Introduction

Protein function can be influenced by the chemical modification of constituent amino acid side chains. These PTMs may occur by default or in response to environmental stimuli whereby the cell integrates signalling cues to elicit an effect. Consequently, the stability, structure, cellular distribution and activity of a protein can be altered by these chemical groups (Knorre, Kudryashova and Godovikova, 2009). In this chapter, I describe the results of an investigation into the effects of the relatively understudied PTM citrullination on the regulation of DNA methylation in mESCs.

Citrullination describes the post-translational conversion of protein arginine residues to citrullines. Biochemically this equates to the loss of a positive charge and increased hydrophobicity. The process is mediated by the highly conserved peptidylarginine deiminases (PADI 1-4 and 6) in a Ca^{2+} dependent manner (Vossenaar *et al.*, 2003; Wang and Wang, 2013).

Whilst PADIs have physiological roles in the adult, they are also a focus of interest in inflammation and autoimmune diseases, particularly rheumatoid arthritis, in which there is an association with deregulated activity of the enzymes and the generation of autoantigens. This precipitates an attack on structural proteins in the synovium and inflammation and degradation of the joints (Reviewed Valesini *et al.*, 2015).

The PADIs have been shown to regulate gene expression. For example, PADI2 is recruited to estrogen receptor α binding sites in response to 17β-estradiol (E2) application in MCF-7 cells. Loaded here, PADI2 citrullinates histone 3 at arginine 26 (H3R26) to facilitate the opening of chromatin and subsequently transcription (Zhang *et al.*, 2012). Furthermore, the enrichment of PADI4 at transcriptional start sites in MCF-7 cells was shown to positively correlate with expression and align with other factors associated with activation, such as E2F1 and RNA polymerase II. In this study, the target of citrullination was identified as the Elk-1 transcription factor, which led to its increased phosphorylation. This is a necessary step to promote an interaction with the histone acetyltransferase p300 and activation of gene expression at *c-Fos* promoters (Zhang *et al.*, 2011). In a different study, the specific citrullination of H3R8 (H3R8Cit) by PADI4 was shown to be antagonistic to HP1α binding, resulting in the expression of ERVs, satellite repeats and some immune response genes in the MCF-7 cancer cell line and Jurkat T cells. This relationship was also observed to an extent in peripheral blood

mononuclear cells from multiple sclerosis patients (MS) (Sharma *et al.*, 2012). As a consequence of the effects on gene regulation, deregulated PADI activity has an association with cancer. PADI2 was notably significantly overexpressed in multiple breast cancer cell lines relative to healthy luminal tissue (Mackay *et al.*, 2009). It has also proven to be an effective target in supressing the growth of xenografted MCF10DCIS tumorigenesis models in mice (McElwee *et al.*, 2012).

More large-scale chromatin effects are induced by citrullination and this is particularly the case with PADI4 mediated activity. In neutrophils, H3/H4Cit promotes expulsion of chromatin, termed neutrophil extracellular traps (NETs), to ensnare pathogens (Wang *et al.*, 2009). The open chromatin structure in mESCs, an important feature of pluripotency, is dependent on PADI4 which acts on the linker histone H1 to reduce the binding affinity to nucleosomal DNA. Depletion or chemical inhibition of PADI4 in these cells reduces the expression of the pluripotency factors *Nanog* and *Tcl1*, at whose promoter and enhancer regions H3Cit is typically present. This also leads to upregulation of differentiation markers such as *Prickle* and *Wnt8a*. Conversely, PADI4 overexpression induces pluripotency related factors including *Tcl1, Tcfap2c* and *Kit* (Christophorou *et al.*, 2014). Accordingly, PADI4 knockdown or inhibition significantly impairs the reprogramming of a neural stem cell line to a pluripotent state and *Padi4*-knockout mice are born at a lower than expected Mendelian ratio (Li *et al.*, 2010; Christophorou *et al.*, 2014).

PADI activity is also detected throughout pre-implantation development: H3Cit26 is likely derived from the oocyte and is enriched in nuclei at the 2-cell stage; H4Cit3 is predominantly cytoplasmic, aligning with chromatin at metaphase and H3Cit 2/8/17 is present in the nucleus during interphase (Kan *et al.*, 2012). The culture of zygotes in the pan-PADI inhibitor Cl-amidine arrests development by the 2-4 cell stage and promotes histone hypoacetylation, suggesting citrullination has a function in this developmental window (Kan *et al.*, 2012). The knockdown of *Padi1* recapitulates these results and negatively affects gene expression, suggesting that this is the dominant isoform in pre-implantation development. Indeed, it is expressed in nuclei of oocytes and from the 2-cell stage onward. This contrasts with PADI2 and -3 which exhibit weak signals in the oocyte and are largely excluded from the nucleus thereafter (X. Zhang *et al.*, 2016). The localisation of PADI4 meanwhile is subject to debate with use of different antibodies and controls, regardless of it being unique amongst the PADIs in having a nuclear localisation signal (Brahmajosyula and Miyake, 2013; X. Zhang *et al.*,

2016). Although PADI6 lacks enzymatic activity, deleterious homozygous or compound heterozygous mutations are associated with infertility, likely owing to a failure to maintain oocyte cytoplasmic lattices that contain a store of ribosomes. This manifests as dysregulated protein synthesis during the 2-cell stage and developmental arrest (Yurttas *et al.*, 2008; Xu *et al.*, 2016).

The examples described above showcase the effects of citrullination on both histone and non-histone targets in a physiological context and instances when deregulation leads to aberrant cell biological functions. Due to the biochemical effect citrullination imparts, there are conceivable implications on the interaction of the protein modified this way with other proteins. As more PADI substrates are found and the effects investigated, our understanding of the function of citrullination will grow. Identifying novel PADI targets can be facilitated with mass-spectrometry however, the small increase in the mass of a citrullinated peptide by 0.98 Da poses a challenge (Clancy, Weerapana and Thompson, 2016).

A previous report has outlined a positive association between PADI4 and DNMT3A protein levels in cancerous cell lines and MEFs. Manipulating the levels of PADI4 or inhibiting its activity was shown to confer a change in DNA methylation at a number of CpGs across two loci tested (Deplus *et al.*, 2014). Furthermore, citrullination is prevalent in pre-implantation development, as discussed earlier, and this is also a period in which DNA demethylation occurs (Smith *et al.*, 2012). This raises the opportunity to use both serum and 2i media to address whether PADI4 is important for DNA methylation in the context of primed and naïve pluripotency.

In this chapter, I describe the results from an investigation into the influence of PADI4 on DNA methylation. This study was built on preliminary data from a mass-proteomic screen performed by Maria Christophorou in which Dnmt1, Dnmt3b and Uhrf1 were shown to be subject to citrullination by PADI4 (unpublished data). Using CRISPR generated *Padi4*-knockout and stably overexpressing human *PADI4* lines, I aim to outline the influence of PADI4 on global methylation levels in serum-based media and during the transition to a more naïve state with KOSR+2i. In an attempt to uncover the mechanism through which PADI4 may influence DNA methylation, the protein and mRNA of the Dnmts and Uhrf1 were also assessed in the aforementioned lines.

5.2 Results

To elucidate the potential influence of PADI4 on DNA methylation, mESCs are suitable tractable models to achieve this aim. Firstly, as shown in Chapter 3, DNA methylation of mESCs can be modulated by transitioning from serum-based media into alternative tissue culture media, KOSR+2i. Whilst there is evidence of PADI4 activity in mESCs cultured in serum-based media, work on the reprogramming of neural stem cells to iPSCs suggests PADI4 activity is responsive to KOSR + 2i (Christophorou *et al.*, 2014). Furthermore, as 2i culture systems have been shown to recapitulate the ground-state of pluripotency *in vivo* to an extent, a period in which PADI activity is detectable, findings may help elucidate novel roles of citrullination in pre-implantation development (Marks *et al.*, 2012; X. Zhang *et al.*, 2016).

5.2.1 The effect of Padi4KO on DNA demethylation in KOSR+2i

For initial experiments, *Padi4*-null (*Padi4*KO) mESC clones were used and compared to a parental WT E14 line. The *Padi4*KO lines, Clone 1 and 2, were generated by Abigail Wilson using CRISPR. Sequencing revealed Clone 1 has a large deletion between *Padi4* exons 7-11. Clone 2 has exon 9 deleted and a frameshift in the gene. As shown in Figure 5.1, the lines do not have detectable PADI4 protein. The first 5 days of culture in KOSR+2i were the primary focus to investigate the influence of PADI4 as this represents the most dynamic period of demethylation in the KOSR+2i culture system, as presented in Figure 3.1. Daily collections were taken of DNA, protein and mRNA. The first step was to characterise the effect of *Padi4* knockout on DNA methylation using LC-MS.

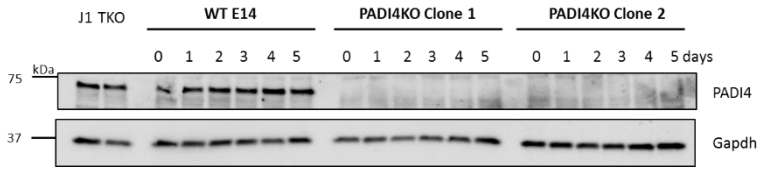


Figure 5.1 – Western blot of whole cell lysates from *Padi4* CRISPR targeted E14 cells and the parental cell line. Gapdh was used as a loading control. Time above each lane indicates days in KOSR+2i culture.

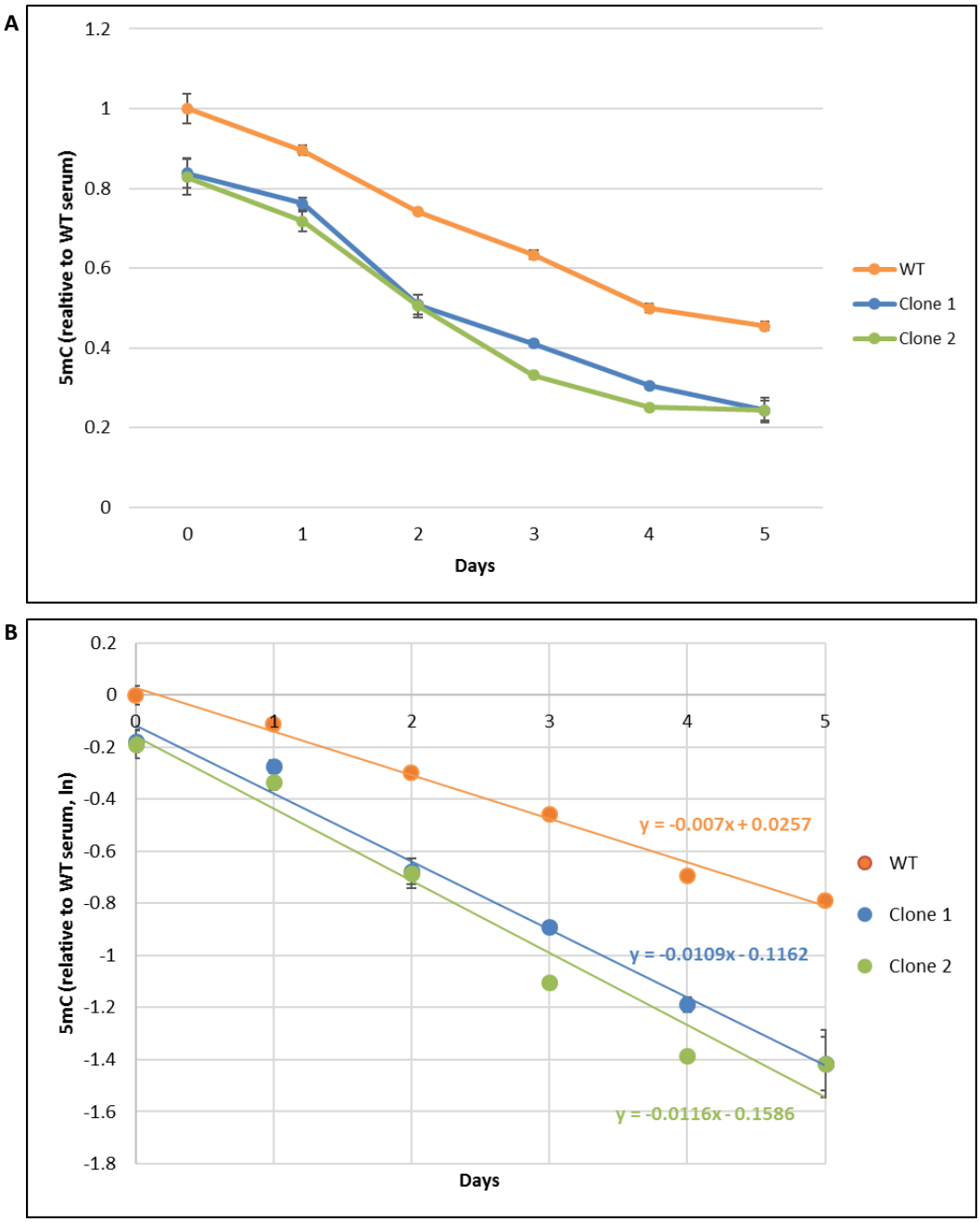


Figure 5.2 – A. DNA methylation levels of WT and *Padi4*KO E14 mESCs cultured in KOSR + 2i for the days indicated. Values are presented as relative methylation levels to WT E14 serum cultured cells at 0 days. **B**. The relative values from days 0-5 of culture presented as natural logs. The slope of the line indicates the rate of demethylation, assuming first order rate kinetics. Error bars represent the standard deviation of the mean from three technical replicates. (**= 0.001<*p*<0.05, *** = *p* ≤0.00001)

Values presented in Figure 5.2 A are plotted relative to the average value of the WT line cultured in full-serum media, this is 0 days. As expected, the WT control underwent changes within 1 day of transitioning to KOSR+2i, and methylation was lower at later time points assayed. The initial global methylation level of both *Padi4*KO clones was 17% lower than that of the equivalent WT line; it required between 1 and 2 days of culture in KOSR+2i for the WT line to be comparably methylated. At all stages thereafter, the knockout clones remained more significantly demethylated than the WT counterpart.

Taking the natural log of the relative values and plotting against time enables the rate of demethylation to be determined in KOSR + 2i. Assuming first order kinetics, this will be equivalent to the slope of the line. As presented in Figure 5.2 B, the rates of the *Padi4*KO clones are relatively faster than that observed in the WT line, differences which are statistically significant (Clones 1 and 2: $p \leq 0.001$). Meanwhile, the *Padi4*KO clones perform similarly to each other ($p > 0.05$).

Collectively, this data demonstrates that PADI4 is important for sustaining DNA methylation levels in mESCs maintained in serum media and for counteracting the mechanisms that promote demethylation in KOSR+2i. As discussed earlier, PADI4 has been shown to influence both transcription (Christophorou *et al.*, 2014) and in the context of cancer cells and MEFs, DNMT3A protein stability (Deplus *et al.*, 2014). Therefore, in order to identify the mechanism through which PADI4 modulates DNA methylation in mESCs shown above, protein and mRNA levels of DNA methylation regulators were determined in western blots and RT-qPCRs, respectively.

When maintained in serum media, protein levels of the *de novo* transferases, Dnmt3a and -b are comparable to the WT line, as is the level of the obligate co-factor of Dnmt1, Uhrf1. There is however appreciably less Dnmt1 protein. *Dazl* is a methylation sensitive gene and its expression is influenced directly by its methylation status. Levels can therefore be used as an indicator of global methylation (Hackett *et al.*, 2012; Walter *et al.*, 2016). Despite the reduction in Dnmt1 protein with *Padi4*KO, Dazl levels are comparable between the WT and *Padi4*KO clones. More differences emerge when the parental line and the knockout clones are transitioned into KOSR + 2i (Figure 5.3).

The downregulation of Dnmt3a protein proceeds at a similar rate in all lines. However, Dnmt3b declines faster and levels in *Padi4*KO clone 2 after 1 day in KOSR+2i culture are comparable to those attained by the WT line in 3 days of culture. Levels of Uhrf1 protein in

both knockout clones reach a nadir earlier than the WT line. Whilst clone specific effects on Dnmt1 are seen, appreciably lower levels are attained earlier in the KOs with KOSR+2i culture also (Figure 5.3).

Furthermore, Dazl protein is detected after 3 days in KOSR+2i culture whereas it is present in the WT line at 5 days. This supports the LC-MS data (see Figure 5.2) which suggests the *Padi4*KO clones demethylate at a faster rate. Noticeably, there is undetectable citrullination, as measured by H3 citrullination, in the *Padi4*KO clones and the WT line too.

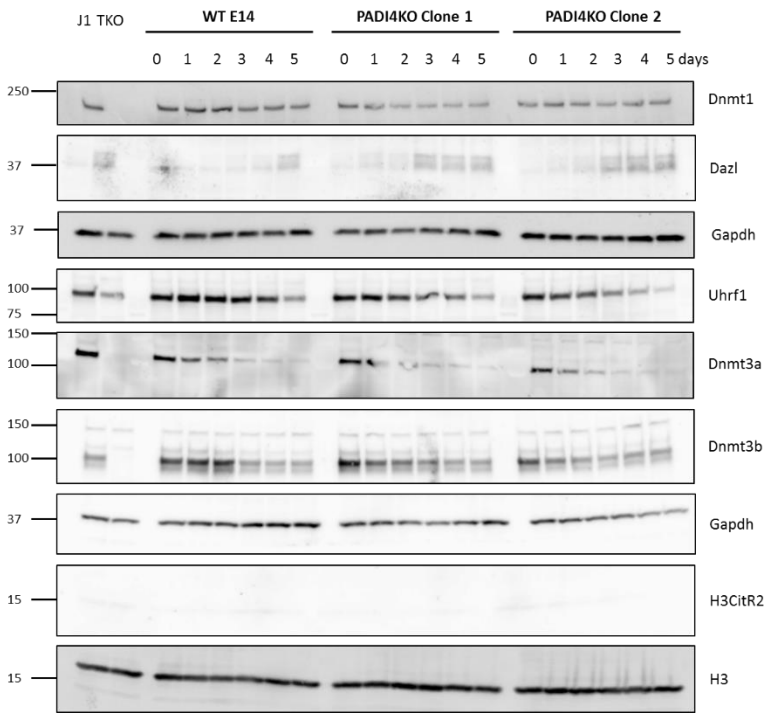


Figure 5.3 – Western blot of whole cell lysates from *Padi4* CRISPR targeted E14 cells and the parental cell line. Gapdh and histone 3 (H3) were used as loading controls. Time above each lane indicates days in KOSR+2i culture. Lysates from J1 and TKO cells have been presented to demonstrate the specificity of the Dnmt antibodies. The TKO lysate also serves as a positive control for methylation sensitive Dazl expression.

To determine whether the differences in the protein levels of the DNA methylation mediators are due to post-translational mechanisms or can be accounted for by direct effects on transcription, RT-qPCR was performed on samples harvested in parallel with protein (Figure 5.4). In serum alone, there is increased mRNA for all these genes in at least one knockout clone. Most strikingly is the almost 2-fold higher *Dnmt3b* mRNA in *Padi4*KO clone 2 however,

with the evidence presented above, this does not appear to have implications on the protein level. Expression thereafter is comparable between the *Padi4*KO clones, except at 3 days, and the WT line. Whereas there is a highly significant difference ($p \leq 0.001$) between *Dnmt3b* expression in *Padi4*KO clone 2 within serum culture (0 days) and at the end point of 5 days, this is not the case for the WT line nor *Padi4*KO clone 1. Therefore, the effects of KOSR+2i overcome the initial influence of *Padi4* absence.

In contrast, *Dnmt3a* mRNA is more significantly ($p \leq 0.05$) abundant in *Padi4*KO clone 1 in serum and remains as such in at least one knockout clone throughout KOSR+2i culture. Comparing the high mRNA levels to protein suggests there is again no correlation. Throughout KOSR+2i culture, *Dnmt3a* levels between the WT line and a knockout clone differ by 40% at the most, as seen at the 4-day time point. After day 1 however, as mRNA in KOSR+2i is significantly less abundant than in serum ($p \leq 0.05$) and remains low, it may be unexpected for a noticeable effect on protein to be seen despite the differences in mRNA.

The overall trend in expression of *Dnmt3l* is similar to *Dnmt3a* in that levels in the knockout clones are initially higher and this remains so in at least one clone at the majority of time points in KOSR+2i. At all stages from day 1 however, levels remain significantly ($p \leq 0.05$) lower than serum culture.

Regarding the components of the maintenance machinery, there is also a deviation between transcript level and protein. Levels of *Dnmt1* mRNA are comparable between lines across the time points except in serum and at day 1 where *Padi4*KO clone 2 expresses 2-fold more, a statistically significant amount ($p \leq 0.05$). This contrasts to protein which is initially lower in both knockout clones. By day 5, the mRNA levels between the three lines are comparable. With respect to the *Padi4*KO clones, this represents a significant ($p \leq 0.05$) decline in *Dnmt1* transcription when compared to expression in serum-based media. Fluctuations in *Uhrf1* expression are detected throughout culture in KOSR+2i in all lines at most time points, with differences reaching statistical significance ($p \leq 0.05$) between the WT and knockout clones on occasion. At the latest time point, *Uhrf1* expression in *Padi4*KO clone 2 is comparable to the WT E14 line.

A direct correlation is observed between Dazl mRNA and protein. Both *Padi4*KO clones behave similarly and levels remain significantly ($p \leq 0.05$) more abundant than the WT line at all time points assayed, including serum. Although *Dazl* is responsive to demethylation, it will also be subject to transcriptional regulation and therefore this alone is an insufficient

indicator of methylation. The *Padi4*KO clones may be more susceptible to demethylation yet the transcription of *Dazl* may be sensitive due to additional signalling mechanisms. Nevertheless, there is a positive correlation between the LC-MS data (see Figure 5.2) and Dazl levels (See Figures 5.3 and 5.4)

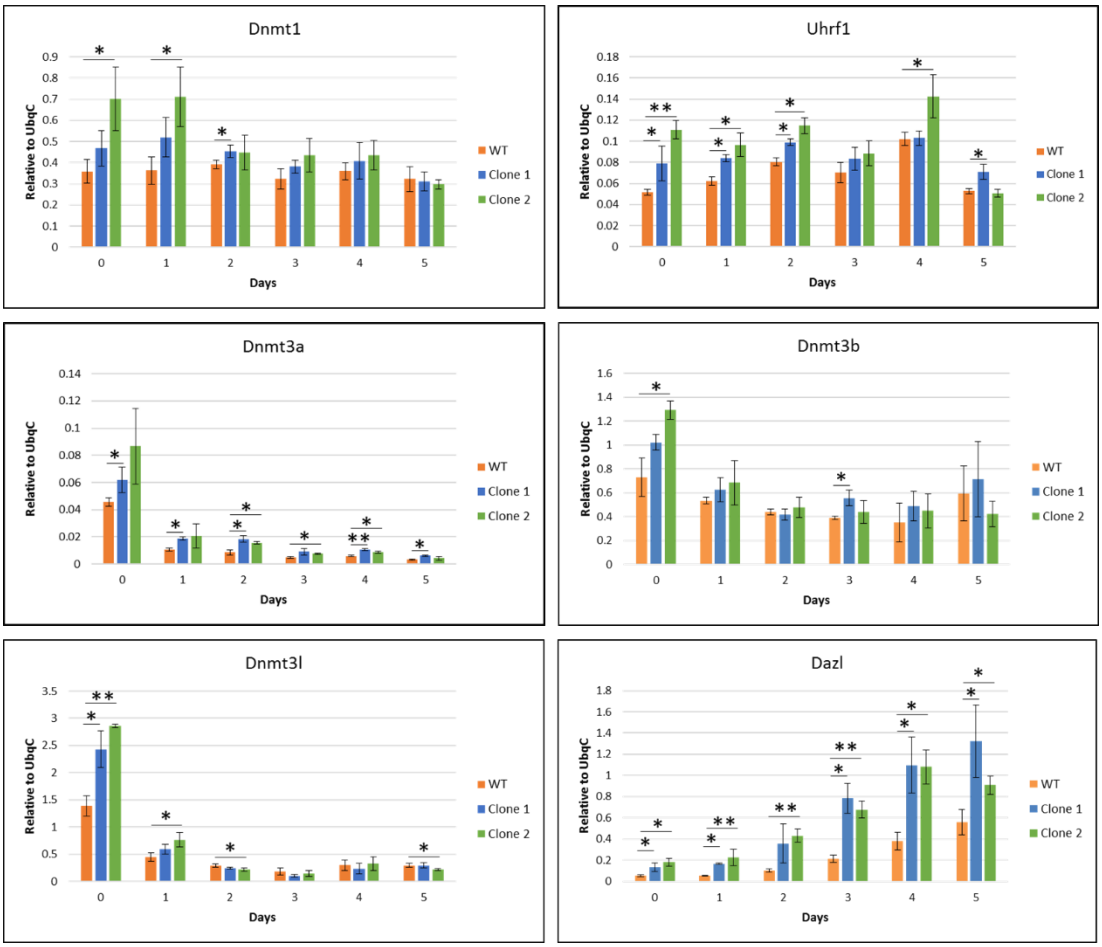


Figure 5.4 – RT-qPCR analysis of DNA methylation mediators in WT E14 cells and two *Padi4*KO clones cultured in KOSR + 2i at the indicated time in days. *Dazl* expression has been determined to indicate the extent of DNA demethylation. Values are presented relative to *UbqC* expression. Error bars represent the standard deviation from three technical replicates. (* = $p \leq 0.05$; **= $0.001 < p < 0.05$).

The above data suggests that PADI4 influences the stability of the DNA methylation mediators examined. Whilst there are some transcriptional changes upon *Padi4*KO, the extent of which for *Dnmt1* and *Dnmt3b* in particular is lessened with culture in KOSR+2i, the western blots demonstrate the effects of PADI4 on protein regulation are maintained in this medium.

5.2.2 The effect of PADI inhibition on DNA demethylation in KOSR+2i

To account for potential compensatory effects from other PADI enzymes expressed in mESCs, such as PADI2 (see Figure 5.5), the level of DNA methylation in KOSR+2i was determined in the presence of a pan-PADI inhibitor.

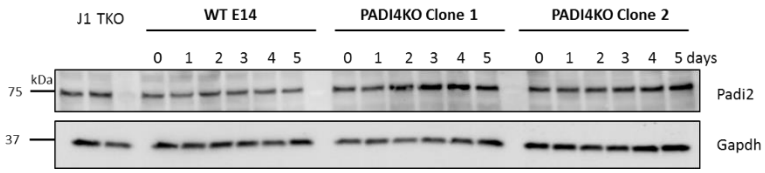


Figure 5.5 – Western blot of whole cell lysates from *Padi4*KO E14 cells and the parental cell line. Gapdh was used as a loading control. Time above each lane indicates days in KOSR+2i culture.

Culturing a WT E14 line in KOSR+2i with the inhibitor promoted demethylation by a further 10% compared to KOSR+2i alone by 2 days, a statistically significant ($p \leq 0.00001$) difference (Figure 5.6). The data from this experiment supports the notion that PADI4 delays demethylation.

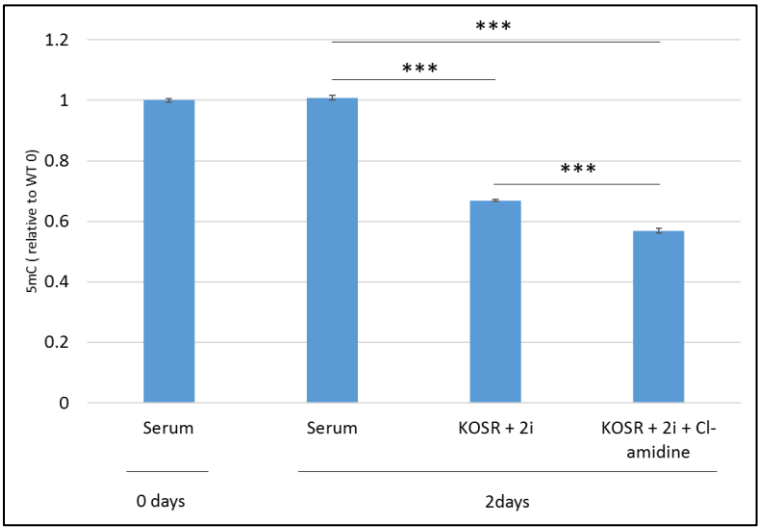


Figure 5.6 – DNA methylation levels of WT mESCs cultured in KOSR + 2i with or without a pan-PADI inhibitor (Cl-amidine) for the days indicated. Values presented as relative levels to WT E14 serum cultured cells at 0 days**.** Error bars represent the standard deviation of the mean from three technical replicates. (*** = $p \leq 0.00001$)

5.2.3 The effect of PADI4 overexpression on DNA demethylation in KOSR+2i

A set of converse experiments were then performed to determine the implications of *PADI4* overexpression. The cell lines used throughout were of a J1 background stably transfected with a vector that overexpresses human *PADI4*, as driven by a CAG promoter. A line expressing an empty vector served as the control line (Figure 5.7).

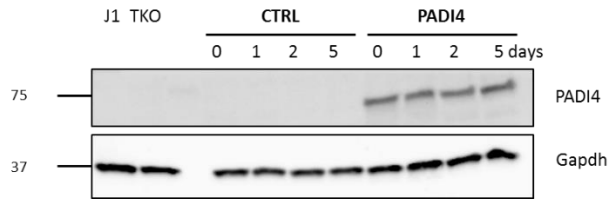


Figure 5.7 – Western blot of whole cell lysates from human *PADI4* overexpressing (PADI4) and empty vector transfected (CTRL) J1 mESCs. Gapdh was used as a loading control. Time above each lane indicates days in KOSR+2i culture.

Assaying for global DNA methylation levels shows that *PADI4* overexpression does not impart an effect in serum. Initially there is no effect detectable in KOSR+2i either however, a small yet statistically significant (p<0.05) divergence at days 2 and 5 is seen, with the PADI4 line being 7% and 8% more methylated at these time points (Figure 5.8). This complements the experiments with the *Padi4*KO lines which suggested PADI4 counteracts DNA demethylation.

In contrast to the *Padi4*KO lines, the *PADI4* overexpressing cells displayed similar levels of Dnmt1 protein in serum media and throughout KOSR+2i culture as the control line. Furthermore, Uhrf1, Dnmt3a and -b and Dazl protein levels were also comparable between the lines. The effects of KOSR+2i on PADI4 activity, using H3 citrullination as an indicator, were also subtle yet insufficient to stimulate the endogenous PADI4 in the control line (Figure 5.9).

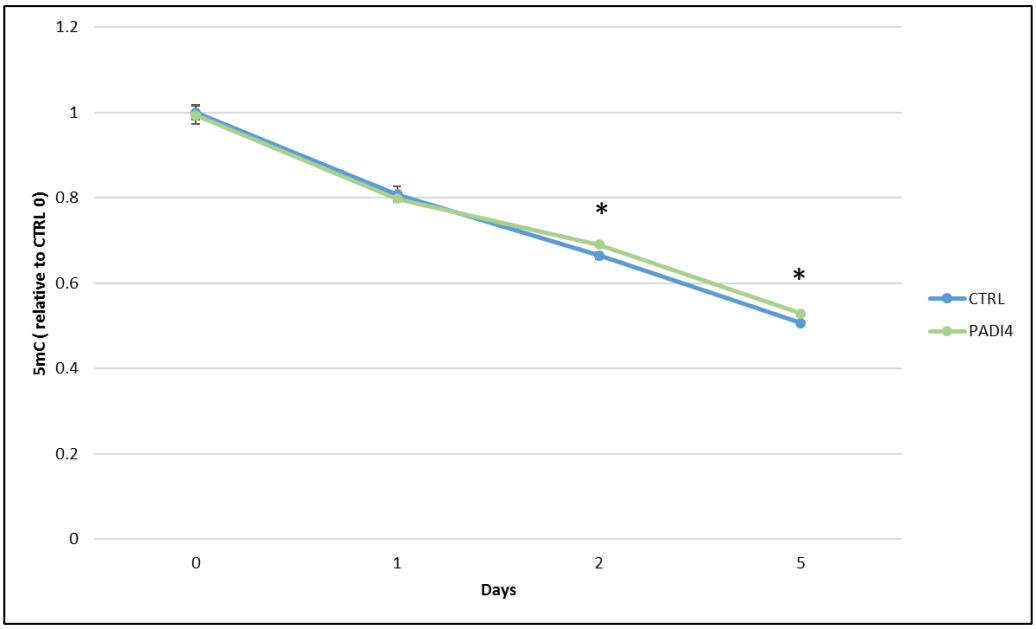


Figure 5.8 – DNA methylation levels of empty vector control (CTRL) and human *PADI4* overexpressing (PADI4) J1 mESCs cultured in KOSR + 2i for the days indicated. Values presented as relative levels to CTRL serum cultured cells at 0 days. Error bars represent the standard deviation of the mean from three technical replicates. (* = $p \leq 0.05$).

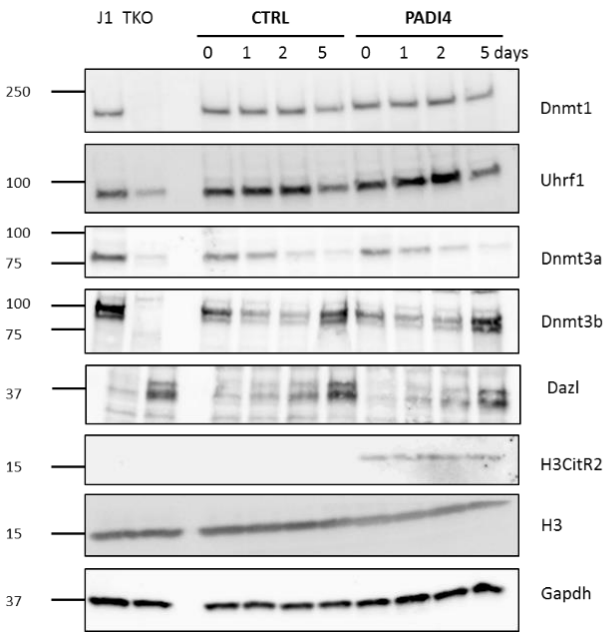


Figure 5.9 – *Figure legend on next page*

Figure on previous page **Figure 5.9** – Western blot of whole cell lysates from human *PADI4* overexpressing (PADI4) and empty vector transfected (CTRL) J1 mESCs. Gapdh was used as a loading control. Lysates from J1 and TKO cells are presented to demonstrate the specificity of the Dnmt antibodies. The TKO lysate also serves as a positive control for methylation sensitive Dazl expression. Time above each lane indicates days in KOSR+2i culture.

Analyses of the mRNA show an inverse relationship between PADI4 and the expression of the *de novo* transferases and co-factor *Dnmt3l* in serum cultured cells. Whilst *Padi4*KO lines have statistically significantly higher levels of these transcripts, *PADI4* overexpression has the opposite effect ($p \leq 0.05$). The changes that KOSR+2i imposes on transcription overcomes the influence of *PADI4* overexpression as levels of genes tested are not significantly different between the control and PADI4 line in this medium, including in the rebound of *Dnmt3l* and -*b* at day 5. An exception is at 2 days for *Dnmt3l* however, both lines have lower levels relative to serum media-maintained cells (Figure 5.10). Although *Dnmt3a* is expressed 2-fold higher and *Dnmt3b* 40% higher in the control cell line in serum-based media, this has no bearing on protein levels (Figures 5.9 and 5.10).

Transcriptional regulation of *Dnmt1* and *Uhrf1* is also comparable between the lines. Fluctuations in the expression of these genes are detected throughout KOSR+2i culture but this does not reach statistical significance ($p > 0.05$) except at day 5 for Uhrf1, which does not have implications on protein abundance. Meanwhile, *Dazl* increases similarly in the lines throughout culture in KOSR+2i and an initial statistically significant difference ($p \leq 0.05$) emerges at day 1 (Figure 5.10).

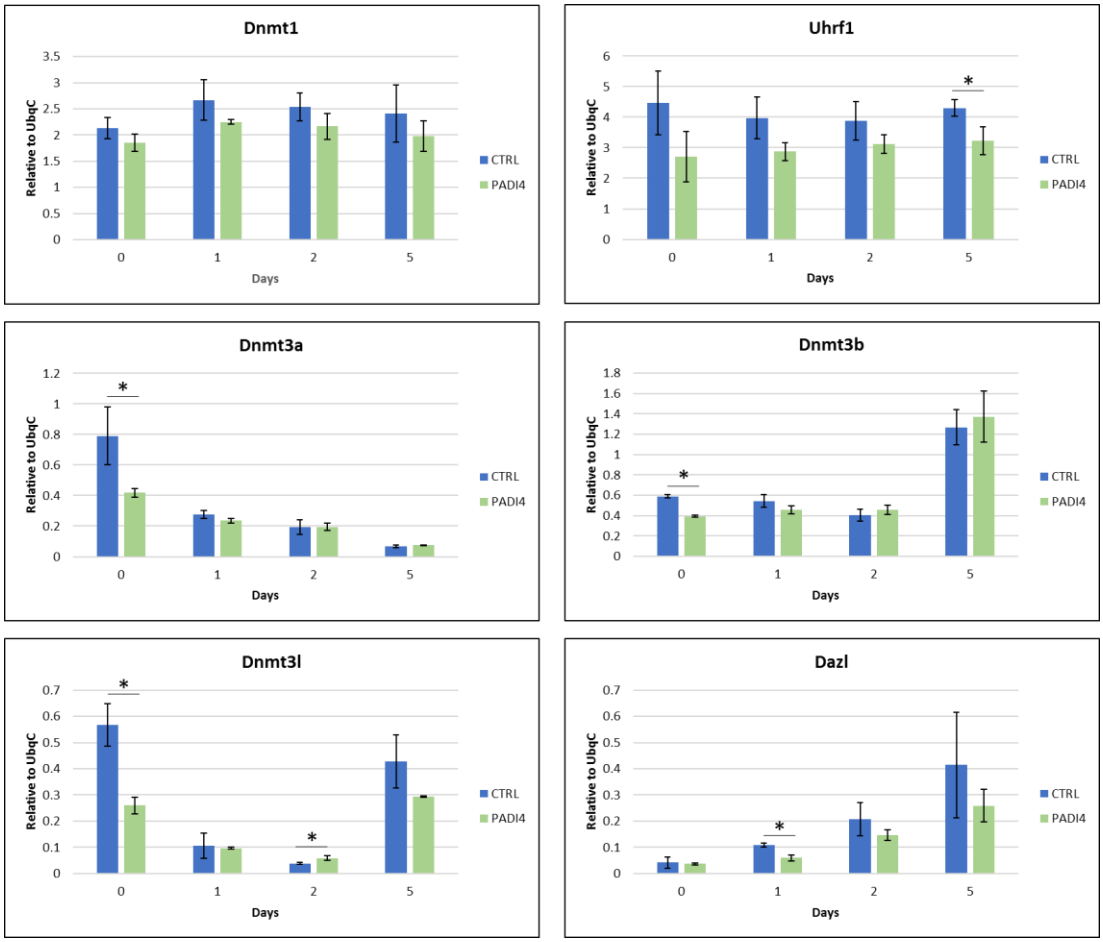


Figure 5.10 – RT-qPCR analysis of DNA methylation mediators in human *PADI4* overexpressing (PADI4) and empty vector transfected (CTRL) J1 mESCs. Time indicates days in KOSR + 2i culture. Values are presented relative to *UbqC* expression. Error bars represent the standard deviation from three technical replicates. (* = $p \leq 0.05$).

5.3 Discussion

In this chapter, the relationship between PADI4 and DNA methylation was investigated. The potential for PADI4 to disrupt DNA methylation could be inferred from the antagonistic effect of citrullination on HP1 binding. As various isoforms of HP1 have been shown to interact with and stimulate Dnmt1 activity, hypomethylation could be promoted indirectly through a destabilised HP1-Dnmt1 complex (Fuks *et al.*, 2003; Smallwood *et al.*, 2007). PADI4 may also influence DNA methylation through the modification of H3R2. This histone residue is contacted extensively through intermolecular interactions by the PHD of Uhrf1. Substitution of the H3R2 residue for a non-polar alanine residue was shown to significantly impair binding affinity, as quantified by isothermal titration calorimetry (ITC). Although a negative effect was measured at specific loci, the global impact is nevertheless likely to be slight as H3R2 is required for euchromatic targeting only (Rajakumara *et al.*, 2011). Pre-implantation development, a period in which epigenetic marks are reprogrammed, is also associated with PADI activity (Kan *et al.*, 2012; X. Zhang *et al.*, 2016). However, despite these aforementioned observations, data from *Padi4*KO and *PADI4* overexpressed mESC lines suggests citrullination facilitates DNA methylation.

The specific genomic regions that are hypomethylated in the *Padi4*KO lines can be identified with bisulfite sequencing, which determines methylation at the resolution of individual CpGs (Li and Tollefsbol, 2011). The biological significance of demethylated repeat elements, which may otherwise be supressed by DNA methylation (Walter *et al.*, 2016), could potentially be greater than demethylated genes, given the general uncoupling of DNA methylation and gene transcription (Hackett *et al.*, 2012).

5.3.1 Padi4 knockout promotes DNA hypomethylation

Possible explanations for the variable demethylation rate between *Padi4*KO and WT E14s include an altered cell-doubling time, reduced protein stability of DNA methylation regulators and impaired Dnmt recruitment, each of which are addressed below.

In initial observations, *Padi4*KO cell counts with a haemocytometer at 2 and 4 days in KOSR+2i culture were ~50% more than WT cells, despite equal numbers of all cell lines being seeded two days prior. As impaired maintenance is the major mechanism that accounts for hypomethylation in 2i (von Meyenn *et al.*, 2016), increased cell proliferation will dilute 5mC more quickly. Furthermore, this effect could also account for the relatively lower global

methylation levels of the serum maintained *Padi4*KO lines. It was previously shown that by arresting hESCs in pro-metaphase with nocodazole, the proportion of symmetrical CpG dinucleotides in a fully methylated configuration increases. These additional sites are restored from hemi-methylated regions by mechanisms that act on DNA after replication-coupled processes. Similarly, when cell division of the highly proliferative HCT116 colon cancer cell line was arrested, there was a concurrent increase in global DNA methylation. Therefore, if a larger proportion of cells are in S phase at a given time, methylation would be notably lower as post-replicative DNA methylation would not have yet occurred (Charlton *et al.*, 2018). The implications of these findings on DNA methylation in the *Padi4*KO lines can be tested experimentally by similarly inhibiting cell replication with nocodazole. As a result, I hypothesise that global methylation levels in the *Padi4*KO lines would subsequently be comparable to the WT E14 line in serum-based media. Initially, cell proliferation should also be investigated using a more robust method. Flow cytometry could be used to quantify nucleoside analogue incorporation, which would be proportional to the number of S phases the population completes, or to measure the fluorescence of cytoplasmic dyes that will become diluted with each cell division (Romar, Kupper and Divito, 2016).

An association between citrullination and cell proliferation has been observed previously in keratinocytes. Psoriasis is characterised by hyperproliferative keratinocytes with diminished levels of citrullinated keratin K1, a process which is typically mediated by PADI1. Whether PADI1 activity directly regulates the proliferation of these cells remains to be explored (Ishida-Yamamoto *et al.*, 2000; Vossenaar *et al.*, 2003). Furthermore, the LSK fraction of haematopoietic cells in *Padi4* null mice proliferate more rapidly than in WT mice. The authors draw parallels with this phenotype and an elevated c-myc expression, which promotes cell proliferation. This was suggested to occur mechanistically by a failure of PADI4 to facilitate the recruitment of transcriptionally repressive HDACs to the c-myc promoter (Nakashima *et al.*, 2013).

An increase in cell proliferation may also account for the effect on Dnmt3b protein levels, which is regulated at the level of transcription in KOSR+2i. As seen in Figures 5.3 and 5.4, Dnmt3b protein declines more quickly in the *Padi4*KO lines compared to WT however, *Dnmt3b* mRNA levels are comparable throughout KOSR+2i culture. Whilst Uhrf1 protein in the *Padi4*KO clones is lower at early time points in KOSR+2i compared to WT E14s, the transcript levels are marginally higher, albeit this is a statistically significance difference. The

reduced Uhrf1 protein would further exacerbate demethylation as the maintenance of 5mC would occur less efficiently at equivalent time points as a WT line. As *Uhrf1* mRNA on the whole is not significantly lower in the *Padi4*KO lines, the post-translation regulation of Uhrf1 appears to be more acute. Furthermore, Dnmt1 protein is initially lower in *Padi4*KO lines in serum-based media and it declines further in KOSR+2i, despite a comparable effect on transcription (see Figures 5.3 and 5.4). Collectively, this suggests that PADI4 may also influence protein stability.

There has been a report that PADI4 specifically regulates DNMT3A protein levels in multiple cell lines (Deplus *et al.*, 2014). In this study, blots of endogenous IPs performed in U2OS cells show that the total fraction of DNMT3A associated with PADI4 is modest however, when both proteins are overexpressed in a HEK-293T line, DNMT3A stability increases with WT PADI4. Conversely, there is no effect on DNMT3A if co-transfected with a catalytically impaired *PADI4* mutant. Furthermore, endogenous DNMT3A protein was upregulated in U2OS cells upon the transient transfection of WT *PADI4*. This led to an increase in the methylation at some CpGs of a DZDT repetitive element. Upon *PADI4* knockdown, methylation was reduced at the *p21* proximal promoter, at which PADI4 has been shown to localise (Deplus *et al.*, 2014).

A pulse-chase assay, as used to illustrate the role of PADI4 in the study above, would also be suitable to investigate the effects of citrullination on Dnmt and Uhrf1 stability in mESCs (Deplus *et al.*, 2014). For example, cells could be transfected with a vector expressing a tagged protein of interest, such as Dnmt1, along with WT or a catalytic mutant *Padi4* before standard pulse-chase experiment protocols are followed. It could be expected that the WT *Padi4* may delay the rate at which Dnmt1 is degraded. In addition, *Padi4*KO mESCs rescued with a stably expressed WT or a catalytic mutant *Padi4* would also provide a useful tool to address the role of citrullination in protein stability. Using these lines, western blots could be performed to assess whether protein levels of the DNA methylation mediators are comparable to the WT line in serum and throughout KOSR+2i culture. In WT *Padi4* rescue lines, protein levels of the Dnmts and Uhrf1 would be expected to be comparable to a WT E14 line in serum media and subsequently in KOSR+2i. Rescue lines would also address whether the effect is directly related to the absence of *Padi4*. As the two knockout clones used in this set of experiments performed similarly in response to KOSR+2i, it would be surprising if off-target effects accounted for any of the observations. However, as the

*Padi4*KO lines may adapt to the absence of *Padi4* after a prolonged time in culture, rescue may not fully reverse the effects on DNA methylation.

The mechanisms presented above as the potential routes through which PADI4 may influence DNA methylation are not mutually exclusive. In addition, PADI4 may also be important in regulating the association of the DNA methylation machinery to chromatin by facilitating protein-protein interactions between histones or co-factors. To address this, IF would be suitable to visualise the localisation of the components. This could be complemented with biochemical techniques to trap Dnmts on chromatin, using 5-aza-CdR for example. Fractions of the chromatin and nucleoplasm from WT and *Padi4*KO lines can then be prepared and the levels of the Dnmts in each probed for in a western blot. If Dnmt1 levels are lower on the chromatin in the *Padi4*KO line, Uhrf1 and Pcna would be the primary contenders for assessing whether Dnmt1 recruitment is impaired. One possible experiment could be an IP to quantify the extent of interaction between Dnmt1 and these co-factors in WT and *Padi4*KO lines.

Another possible scenario related to protein interactions involves the association between histone citrullination and an open chromatin conformation (Christophorou *et al.*, 2014). As has been suggested with TopII-α inhibition, if chromatin is not presented in a favourable conformation for the activity of Dnmts, then DNA methylation will be negatively impacted (Lu *et al.*, 2015). Likewise, histones have also been shown to be an impediment to DNA methylation maintenance. Whilst *Uhrf1*-null models are embryonic lethal, *Uhrf1* mutants unable to interact with H3K9me2/3 cause on average a 10% reduction in methylation levels, particularly at sequences closest to histones. Mice carrying this mutation are however viable. The implication therefore is that H3K9me2/3 enables Uhrf1 to recruit Dnmt1 to sequences proximal to histones (Zhao *et al.*, 2016). In this respect, PADI4 may be important to arrange chromatin in a conformation that facilitates Dnmt1 association. This possible role would be reflected in experiments discussed above that aim to address whether PADI4 in general is important for facilitating Dnmt1 and Uhrf1 recruitment to chromatin.

The conclusions from the *Padi4*KO lines suggest there is no compensation by other PADI enzymes to mitigate the effects on DNA methylation. Furthermore, global 5mC levels in KOSR+2i were 10% lower at 2 days with pan-PADI inhibition, whereas the *Padi4*KO lines had around 40% less methylation than the WT line at 48 hours (Figures 5.2 and 5.6). Whilst this supports the notion that citrullination positively influences DNA methylation, the discrepancy

between *Padi4*KO and PADI inhibition remains to be investigated. This may be due to the effects on the protein levels of the DNA methylation machinery having not yet manifested with acute inhibition. The *Padi4*KO lines meanwhile have been maintained in serum-based media for multiple passages prior to the study commencing. The efficiency of the inhibitor in ablating PADI4 activity may also be less compared to generating a knockout.

However, there does appear to be an increase in the expression of PADI2 upon transferring the *Padi4*KO lines to KOSR+2i media (Figure 5.5). Whether this may have implications on DNA methylation at the latter stages of KOSR+2i culture could be determined by applying the pan-PADI inhibitor when PADI2 is upregulated, such as after 3 days. Examining the methylation levels of the *Padi4*KO lines in KOSR+2i with or without Cl-amidine would also illustrate if additional PADIs compensate. If the *Padi4*KO lines do not demethylate in KOSR+2i further with Cl-amidine compared to an untreated *Padi4*KO line, then it could be concluded that other PADIs do influence this process.

5.3.2 PADI4 overexpression promotes the retention of DNA methylation in KOSR+2i media

Despite the dramatic effects of *Padi4* knockout, *PADI4* overexpression upregulated *Dnmt3a* and *-b* transcription in serum only, but this did not impact protein levels. This suggests that the influence of endogenous PADI4 on DNA methylation is highly regulated in mESCs as there was no striking inverse effect upon introducing exogenous human PADI4, compared to the *Padi4*KO lines. However, at the 2- and 5-day time points, DNA methylation was retained to a statistically higher level in the *PADI4* line. To investigate the mechanism of this retention, analysing the recruitment of the Dnmts to chromatin and the interaction of Dnmt1 with Uhrf1, as discussed above in the context of the *Padi4*KO lines, would be relevant experiments.

5.3.3 Exploring the role of PADI4 in modulating DNA methylation in alternative systems

A previous study aimed to address the influence of PTMs that may be added or removed from Dnmt1 in a 2i system (Habibi *et al.*, 2013). Dnmt1 immunoprecipitated from mESCs cultured in N2B27 + 2i was shown to be marginally more efficient in an in *vitro* catalytic activity assay, compared to Dnmt1 from serum-maintained mESCs. In isolating Dnmt1 however, the interaction between the cofactors required *in vivo* and the influence that the potential disruption of these could have on DNA methylation was not accounted for. The choice of substrate in the assay, poly dI-dC, is an unmethylated synthetic DNA duplex and is

not a true representative of the hemi-methylated DNA that Dnmt1 has an overwhelming preference for (Pedrali-Noy and Weissbach, 1986; Habibi *et al.*, 2013). Whilst biochemical assays certainly have their advantages, it is then important to understand the context of PTMs in a cellular system.

The data above represents steps to address the potential role of PADI4 in regulating DNA methylation. Whilst Dnmt1, Uhrf1 and Dnmt3b were shown to be citrullinated in mESCs, by using *Padi4* knockout lines and *PADI4* overexpressing systems, the implications on DNA methylation outlined above could in fact be due to indirect effects. To deconvolute this, it may be necessary to mutate Dnmt1, -3b and Uhrf1 to render them resistant to citrullination and conversely, behave biochemically similar to citrullinated proteins. To this end, a previous study substituted arginine for lysine and used alanine to mimic the loss of a positive charge that occurs with citrullination (Christophorou *et al.*, 2014). These mutated constructs could then be transfected to restore proteins in corresponding null cell lines. This will be challenging as the screen in which the citrullinated sites were found used only one enzyme, Lys-C, to digest proteins into the prerequisite peptides for identification in mass spectrometry. As this would only yield suitable peptides that account for a particular percentage of a protein of interest, the use of alternative proteases may be necessary to capture all residues and therefore all citrullinated sites.

Notably, I have been unable to present evidence that PADI4 in mESCs is stimulated by KOSR+2i at the defined time points analysed (see Figure 5.2). Evidence of PADI4 activity in mESCs has been shown with the use of ChIP, which when translated to total protein in the context of a western blot may fall below the detection limit (Christophorou *et al.*, 2014). A direct relationship between regions at which PADI4 localises and the methylation status of the underlying DNA sequence may explain the impact on global methylation levels in the absence of PADI4. Locus specific analysis of DNA methylation, for example using bisulfite sequencing, cross referenced with ChIP data could be used address this point. This approach has already been taken in MCF-7 cells and an inverse relationship between PADI4 bound regions and 5mC was noted from ChIP data. In this context, PADI4 was found to be more highly enriched at actively transcribed genes (Zhang *et al.*, 2011) and therefore the recruitment of transcription factors, facilitated by PADI4, may have inhibited DNA methylation. This may therefore suggest that the loss of methylation recorded in the *Padi4*KO mESCs lines occurs at genomic regions that are not enriched for genes, such as at repetitive

elements. Whilst a cell specific effect also needs to be considered, given that MCF-7 cells are a cancer line, there is also the potential for *Padi4* deletion to promote a redistribution of DNA methylation in mESCs. For example, if PADI4 does negatively affect the deposition of DNA methylation at TSSs in mESCs, its deletion may enable 5mC to be added at regions previously refractory to this mechanism. Deconvoluting a direct effect will be difficult however as transcriptional changes that arise upon PADI4 deletion will indirectly influence DNA methylation.

A context in which PADI activity is unambiguously detected is throughout the reprogramming of pre-iPSCs to iPSCs, a system first discussed in the introduction of Chapter 3. This system could be employed to track the relationship between PADI4 activity and DNA methylation, if this epigenetic mark is indeed dynamic throughout reprogramming. The availability of a cell line in which *GFP* is expressed as a reporter of endogenous *Oct4*, as opposed to the *Oct4* initially transduced along with *Klf4* and *c-Myc*, enables the isolation of fully reprogrammed cells. Of interest, PADI inhibition reduces reprogramming efficiency. Whether this is partially contingent on DNA methylation could be resolved by altering the residues of Dnmt1 for example that are substrates for PADI4 so they are resistant to citrullination or to mimic the modification, evoking a constitutive citrullinated state (Silva *et al.*, 2008; Christophorou *et al.*, 2014). The effect on the reprogramming efficiency could then be determined by GFP levels, as a measure of *Oct4* expression. Beyond the influence of PADI enzymes specifically, these experiments may contribute answers to the question as to whether DNA methylation alterations are strictly necessary for reprogramming.

Collectively, this data suggests that citrullination may not influence DNA demethylation in pre-implantation development *per se*, despite its prevalence throughout this window. Nevertheless, *PADI4* overexpression was shown to delay demethylation, albeit only marginally. Therefore, whether PADI4 contributes towards maintaining the methylation status of loci that remain hypermethylated against a background of global reprogramming would be intriguing to follow (Smith *et al.*, 2012). Given the different phenotypes of *Padi4* and *Dnmt* deletion, as addressed above and in the Introduction, the influence of PADI4 may be minimal in this regard (Li, Bestor and Jaenisch, 1992; Okano *et al.*, 1999; Li *et al.*, 2010). As discussed above, PADI4 has been shown to regulate the stability of DNMT3A in cancer lines and MEFs. Whilst the relevance of these findings to mESCs requires thorough investigation, overexpression of *PADI4* was reported in the same study to increase

endogenous DNMT3A protein (Deplus *et al.*, 2014). Notably, this was not evident for this particular protein in mESCs (see Figure 5.9) and therefore tissue specific effects need to be considered. This also provides further impetus to investigate the role of PADI4 in modulating the additional proposed mechanisms, namely the cell cycle or recruitment of Dnmts, which will have implications on DNA methylation.

Chapter 6 – General discussion

6.1 Novel regulatory mechanisms of DNA methylation

Attempting to explain the molecular mechanisms behind an event that occurs *in vivo* is often an intriguing proposition alone. However, as is the case with DNA methylation reprogramming in early development, when the question as to why it is necessary to occur in the first place has not been conclusively answered, this provides further stimulus to resolve the matter. The aim of my project has been to identify and characterise mechanisms that regulate DNA methylation maintenance in mESCs in models of primed and naïve pluripotency. Using serum-based media, a condition that sustains DNA methylation at high levels, and KOSR+2i, which promotes a global loss of 5mC, the differential interacting profile of Uhrf1 has been mapped with BioID, a proximity ligation technique. An attempt was also made to conduct a similar BioID experiment to capture interacting partners of DNMT1 in these different media however, this proved to be less successful. Nevertheless, on one occasion discussed below, the DNMT1 BioID data complemented the Uhrf1 screen. Evaluating mechanisms of DNA methylation regulation also led to the role of citrullination being considered. Preliminary data showed that this PTM acts on Dnmt1, Uhrf1 and Dnmt3b in mESCs.

The main conclusions from these studies are outlined below:

- DNA methylation is rapidly lost in KOSR+2i culture conditions, as has been reported previously for N2B27+2i. Demethylation is initiated in KOSR+2i before the onset of Uhrf1 protein downregulation.
- Uhrf1 forms alternative protein complexes in serum and KOSR+2i media, as determined by BioID and affinity pull-down. Previously known interacting partners were found, validating the experimental approach.
- In the absence of *Padi4*, mESCs display reduced levels of the DNA methylation mediators and are more susceptible to hypomethylation in KOSR+2i media. Therefore, PADI4 modulates DNA methylation by promoting the deposition or maintenance of 5mC.
- PADI4 overexpression has a small but statistically significant effect on delaying DNA demethylation in KOSR+2i media.

6.2 Characterising the interactome of the DNA methylation maintenance machinery

The downregulation of Uhrf1 protein reported in the N2B27 system and in Chapter 3 in KOSR with 2i, undoubtedly contributes to DNA hypomethylation. As outlined in Chapter 1, there are a number of mechanisms that have been shown to regulate Uhrf1 protein stability. It would therefore be sensible to ascertain whether they are relevant to the context of Uhrf1 regulation in 2i media. However, in KOSR+2i, as presented in Chapter 3, DNA demethylation proceeded before an appreciable decrease in Uhrf1 protein was evident. This leaves open the possibility that alternative mechanisms instigate the hypomethylation.

My proteomic analyses have yielded an array of proteins that are potential novel regulators of the DNA methylation maintenance machinery proteins. These adhere to the strictest control in the screen, enrichment over the free BirA*. The initial steps to establish a possible role of a candidate in regulating DNA methylation is to perform knockdowns, followed by quantification of global 5mC levels. Transient transfections of siRNAs would be suitable for assaying levels in serum-based media however, to investigate whether the kinetics of DNA hypomethylation are impacted in the absence of the candidate, a constitutive knockout would be required. To this end, I have cloned *Cas9* into a piggyBac vector and this has been transfected into E14 cells in order to generate knockout lines by transfecting specific guide oligonucleotides.

In the case of Stella, it has since come to light that the role of this protein extends beyond protecting the maternal genome from active demethylation mechanisms. It has now been shown to promote a dissociation between Uhrf1 and Dnmt1, leading to impaired maintenance methylation (Mulholland *et al.*, 2018; Y. Li *et al.*, 2018). This serves as a positive control for the proteomic screens I conducted, suggesting that other proteins revealed to differentially interact may represent additional mechanisms that influence DNA methylation. It also demonstrates that KOSR+2i is a dynamic tool in concert with serum-based media to uncover more details about Stella mediated regulation, providing an alternative to N2B27+2i. Demethylation in these systems will occur against a background of different signalling and transcriptional changes and therefore it will be of interest to dissect the implications of this.

Previous genetic perturbation studies have already revealed that Utf1 influences DNA methylation (Kooistra *et al.*, 2010). As the proteomics screens I conducted have shown Utf1 interacts more favourably with Uhrf1 in serum media, there is the opportunity to demonstrate a biologically relevant context in which a differential interaction leads to

changes in DNA methylation. As the Utf1-Uhrf1 interaction is enriched in serum media, wherein DNA methylation is high, it may appear to be incompatible with the fact that Utf1 is downregulated upon differentiation (Kooistra *et al.*, 2010), which is accompanied by an increase in DNA methylation *in vivo* (Meissner *et al.*, 2008). However, this conclusion would be derived from the assumption that there is a clear division between Utf1 downregulation in differentiation and the re-methylation of DNA. The stages between a naïve state and a committed state are being refined further, with the existence of a formative state being postulated (Smith, 2017). This state of exiting pluripotency but not yet fully committing to differentiation may represent a window in which DNA methylation increases and Utf1 remains expressed. Herein, Utf1 may act instructively in regulating DNA methylation. This is particularly important as the role of DNA methylation in post-implantation development is essential (Li, Bestor and Jaenisch, 1992; Okano *et al.*, 1999). Whilst work is continuing on culture conditions that sustain the proposed formative state, characterising the potential role of Utf1 on Uhrf1 regulation will serve as a proof of principle study.

Of the candidates, Pdk1 is unique in having been described previously as carrying out a PTM, specifically phosphorylation (Bowker-Kinley and Popov, 1999). The concern with Pdk1 however is that its localisation has been defined as mitochondrial on databases online. However, as it is also similarly enriched in serum-based media in the DNMT1 BioID screen, it should not be discounted. As discussed in Chapter 4, other metabolic enzymes such as Pdc, to which Pdk1 binds have been shown to translocate to the nucleus wherein histone acetylation is modulated (Sutendra *et al.*, 2014). As it may also interact with DNMT1 directly, characterising Pdk1 will be facilitated by the array of *in vitro* assays that have been developed to measure the catalytic activity of Dnmt enzymes (Reviewed Poh, Wee and Gao, 2016).

In a similar regard to Pdk1, Apobec3 is associated with Uhrf1 more robustly in serum media. As Apobec3 has been described as a component of the active demethylation pathway (Nabel *et al.*, 2012), it may be expected to associate with Uhrf1 more favourably under conditions of DNA hypomethylation. Furthermore, levels of Apobec3 decline only slightly in KOSR+2i. This may suggest that the signalling changes, rather than the expression changes, that occur upon transitioning into different media are the most influential factors dictating an interaction. Due to the necessity of normalising LFQ values of proteins in serum media however, the fold-difference compared to the average LFQ value in KOSR+2i media did not pass the threshold for Apobec3 to be regarded as a differential interactor. Whether a Uhrf1-

Apobec3 complex instructs DNA methylation in serum and KOSR+2i media can be elucidated by genetically perturbing *Apobec3* and assaying 5mC levels.

In the case of Auts2, an interaction was also found in both culture conditions, in which it was shown to associate with Uhrf1 to comparable extents. Aberrant epigenetic regulation has been observed in neurological conditions (Reviewed Gapp *et al.*, 2014), as discussed in Chapter 4 in the context of Lsh and ICF. Whilst there are a small number of reported studies into Auts2, which were addressed in Chapter 4, focussing on the potential influence of aberrant DNA methylation patterns as a result of Auts2 disruption would be keeping in line with the association of epigenetics and disease.

6.3 Future directions for the interactome screens

If the results from the genetic perturbations suggest that the associations between the candidate regulators and Uhrf1 are functional, and in the case of Pdk1 with DNMT1 also, the next step would be to confirm an interaction. In Chapter 4, I presented evidence from an endogenous Uhrf1 IP however, I was unable to demonstrate an interaction with any of the candidates. A sub stoichiometric or transient interaction were presented as possible explanations for this observation. Several alternative approaches, such as those reviewed in Phizicky and Fields, 1995 and discussed in the in the context of Dnmt1-Uhrf1 interactions in Chapter 1, will be considered to ascertain whether Uhrf1 does indeed interact directly with the candidates. However, if the candidate was shown to associate with Uhrf1 in the BioID screen only, this may suggest it forms a complex with other proteins but this component itself does not bind directly to Uhrf1. In this scenario, immunofluorescence will be a suitable tool to ascertain a relationship with the candidate.

Upon completing the cloning of the BirA* vectors and optimising the necessary procedure to conduct a BioID screen, the report on the development of TurboID was released (Branon *et al.*, 2018). TurboID is an exciting new tool and its application is already proving fruitful in identifying novel interacting partners of signalling mediators (Kim *et al.*, 2019). Furthermore, parallel experiments with BioID demonstrates the superiority of TurboID in detecting a larger number of interacting candidates within a shorter period of time (Zhang *et al.*, 2019). Repeating the DNMT1 interactome screen with the TurboID ligase therefore may prove more relevant than using BioID. For this reason, our understanding of the interacting profile of Uhrf1 would also benefit from repeating with TurboID.

Whilst this project was concerned with the role of Uhrf1 in regulating DNA methylation, a host of other interacting proteins were found in the BioID data that illustrate its diverse roles. As discussed in Chapter 4, chromatin remodellers and pluripotency factors, such as Klf4, were found. This is in line with the role Uhrf1 plays in regulating bivalent domains and acquisition of pluripotency in reprogramming (Kim *et al.*, 2018). Another function of Uhrf1 that was represented in the BioID data is the mediation of the DNA damage response. For example, phosphorylated UHRF1 was shown to interact with BRCA1 at double-stranded break sites on chromatin. Localised here, UHRF1 ubiquitinates RIF1 using its E3 ligase domain and this compromises the ability of RIF1 to interact with 53BP1. As a consequence, this promotes DNA repair through homologous recombination. This is notably less error prone and mutagenic than non-homologous end joining (H. Zhang *et al.*, 2016). In the Uhrf1 BioID screen, both 53bp1 and Rif1 were detected, along with Rnf169 which also positively influences homologous recombination (An *et al.*, 2018). Huwe1, which knockout HeLa models suggest promotes 53BP1 recruitment, was also enriched in the Uhrf1 BioID screen (Mandemaker *et al.*, 2017). Evidence of DNA damage mechanisms could be expected in mESCs as the lack of a G1/S phase checkpoint and basal levels of single stranded breaks and γH2AX have been found in this cell line (Chuykin *et al.*, 2008).

The aforementioned DNA methylation-independent functions, and possibly others not yet defined, may explain the discrepancy in the differential gene expression changes seen previously upon *Uhrf1* knockout in mESCs, compared to *Dnmt1* knockout. Accounting for a 0.05 false discovery rate cut-off, an overlap of only 12.3% of the up-regulated genes and 4.4% of the down-regulated genes was recorded (Kim *et al.*, 2018). The specific knockout of maternal *Uhrf1* also leads to a more severe developmental phenotype. However, it is important to note that the demethylation that accompanies *Uhrf1* knockout is also more severe than *Dnmt1* knockout. For example, ICRs are approximately half as much methylated in the absence of *Dnmt1* compared to controls, whereas they are only a quarter of the level methylated with *Uhrf1* knockout. A similar effect is also seen with IAP elements (Maenohara *et al.*, 2017). With an advancement in tools used for proteomic screens, the route to deconvoluting the methylation-dependent and methylation-independent functions of Uhrf1 would be best served using techniques like BioID and TurboID. Furthermore, consideration should also be given to the impact on the methylation independent functions of Uhrf1 when it is regulated in pre-implantation development, during which DNA hypomethylation is promoted.

6.4 Citrullination as a means to regulate DNA methylation

Having considered the interactomes of Uhrf1, as was the intention for DNMT1, focus then turned to PTMs. This investigation was built on preliminary evidence from a proteomic screen in which Dnmt1, Uhrf1 and Dnmt3b were shown to be chemically modified by PADI4 in mESCs. Whilst none of the PADIs were detected in any samples in the interactome screens I conducted, PADI4 demonstrably influences DNA methylation. Interestingly, when hypothesising how PADI4 may affect DNA methylation in the context of previous papers, as set out in Chapter 5, it was shown to facilitate DNA methylation.

Support from the literature that DNA methylation is influenced by PADI4 can be found in a study in which *PADI4* was genetically perturbed in cancerous and MEF cell lines (Deplus *et al.*, 2014). Importantly, the function of PADI4 mediating DNMT3A stability that was proposed in this report cannot directly account for the effect I observed in mESCs. Whilst there was no effect on Dnmt3a levels in *Padi4*KO mESCs, Dnmt1 protein was appreciably reduced. It is also conceivable that impaired Dnmt1 activity will impose a greater effect on differentiated cells, such as MEFs, in which Dnmt3a and Dnmt3b are lowly expressed (Okano, Xie and Li, 1998). However, comparisons between my results and the aforementioned study would have been facilitated greatly if global methylation analyses had been conducted as the impact on DNA methylation was assayed at two loci only.

Whether an effect on Dnmt1 protein levels also manifests in cancerous tissue would be an interesting avenue to explore. *Padi4* is aberrantly expressed in multiple cancers such as gastric adenocarcinomas, lung adenocarcinomas and hepatocellular carcinomas (Chang *et al.*, 2009). Furthermore, there is a correlation between *Padi2* expression and progression from normal mammary tissue to fully malignant breast carcinomas, and its inhibition delays the growth of tumours in culture and xenografts (McElwee *et al.*, 2012). A similar relationship with tumour progression has also been shown in prostate cancer, in which PADI2 is associated with activating androgen receptor signalling (Wang *et al.*, 2017). Whilst the biological significance of alterations in DNA methylation remains debated (Sproul *et al.*, 2012), elucidating the mechanism that accounts for these changes would nevertheless be a useful contribution, particularly if interventions to restore normal methylation patterns are therapeutically beneficial.

Three non-mutually exclusive and testable hypotheses have been presented to account for the methylation changes in the *Padi4*KO mESCs. Evidently however, the high levels of PADI4

that are found in the multiple cancers discussed above do not ensure cell division is maintained at a comparable rate to normal tissue. Therefore, the exploration of whether PADI4 dependent cell-cycle cycle effects contribute to hypomethylation in cancer will be particularly difficult as PADI4-independent mechanisms that maintain the high proliferation rate in cancer may continue to promote global hypomethylation upon PADI4 inhibition or knockout (Charlton *et al.*, 2018).

6.5 Implications of findings

Across these two defined projects, the data I have gathered has the potential to make Figure 1.8, outlining how DNA methylation maintenance is regulated, even more convoluted yet detailed. The BioID screen offers the chance to identify novel regulators of Uhrf1 that may indirectly influence DNA methylation. By following the steps outlined above, chiefly confirming the interaction and conducting functional assays upon genetic perturbation of the candidate regulators, these will define whether the candidate is taken forward to dissect the mechanism by which it may influence DNA methylation. For the PADI4 project, there is a phenotype, DNA hypomethylation upon *Padi4* knockout, and therefore the next steps will be to uncover the mechanism that accounts for this effect.

In the context of 2i media, these projects have aimed to bridge the gap between cell signalling changes that occur upon transitioning from serum-based media, and the underlying changes to DNA methylation. Whilst 2i media was used as a tool in these projects, it will also be of interest to ascertain whether the conclusions drawn from experiments outlined in this thesis can be related back to *in vivo* development. This will add further details to the picture of events that occurs in pre-implantation development in the context of DNA methylation and inform future studies in which these mechanisms are perturbed to understand the necessity of this epigenetic reprogramming.

Appendix 1- Tables of proteins identified in Uhrf1 BioID screen

Appendix Table 1.1 – The most abundant proteins ranked by average LFQ intensity detected in **A)** BirA* N-terminally tagged Uhrf1 compared to an Empty vector control line in serum based media and **B)** KOSR+ 2i media**; C)** C-terminally tagged Uhrf1 compared to an Empty vector control line in serum based media and **D)** KOSR+ 2i media**; E)** BirA* N-terminally tagged Uhrf1 compared to a free BirA* control line in serum based media and **F)** KOSR+ 2i media**; G)** C-terminally tagged Uhrf1 compared to a free BirA* control line in serum based media and **H)** KOSR+ 2i media

A)

Protein IDs	Gene names	Unique peptides		Coverage (%)		LFQ Intensity	
		Empty	N-BirA* Uhrf1	Empty	N-BirA* Uhrf1	Empty	N-BirA* Uhrf1
Q91ZA3	Pcca	55	57	62.8	64.5	4.97E+09	1.17E+10
Q69ZK6	Jmjd1c	37	123	19.3	48.8	1.23E+08	7.28E+09
E9Q4Z2	Acacb	80	96	37.1	41.1	1.77E+09	6.80E+09
Q8VDF2	Uhrf1	10	52	20.2	52.8	3.27E+07	6.50E+09
E9PVX6	Mki67	3	168	1.3	54.8	6.32E+05	5.13E+09
P63017	Hspa8	37	41	60.9	63.7	1.34E+09	4.67E+09
A2APB8	Tpx2	4	72	4.9	64.5	3.09E+06	3.63E+09
Q6PR54	Rif1	6	79	3.5	33.3	1.10E+07	3.09E+09
Q91ZW3	Smarca5	4	45	7.1	51.7	8.75E+06	2.83E+09
P37913	Lig1	9	47	11.7	48.6	1.40E+07	2.66E+09
Q8C827	Zfp62	1	34	1.7	35.5	1.10E+07	2.59E+09
Q99MD9	Nasp	3	41	3.0	51.8	9.62E+05	2.49E+09
E9Q7G0	Numa1	11	111	6.4	51.5	1.23E+07	2.44E+09
Q62318	Trim28	5	41	6.8	41.7	3.39E+06	2.13E+09
O88491	Nsd1	2	93	0.7	39.1	2.72E+06	2.11E+09
Q61937	Npm1	5	23	28.1	59.7	4.17E+07	1.75E+09
Q6PFD9	Nup98	22	23	16.7	17.3	7.52E+08	1.48E+09
Q6PDQ2	Chd4	3	66	2.7	37.5	4.14E+06	1.40E+09
Q3U9G9	Lbr	16	32	21.7	34.8	7.60E+07	1.18E+09
P09405	Ncl	4	42	9.3	46.0	1.42E+07	1.07E+09
P54276	Msh6	13	48	9.3	33.5	1.44E+07	9.41E+08
Q3URK3	Tet1	2	43	0.8	24.2	3.73E+05	8.02E+08
Q80W00	Ppp1r10	1	37	1.6	46.7	0.00E+00	7.71E+08
P59328	Wdhd1	0	35	0.4	38.0	0.00E+00	6.74E+08
Q01320	Top2a	13	44	11.5	31.5	3.09E+07	6.56E+08
Q9ESU6	Brd4	0	27	0.2	24.4	0.00E+00	5.95E+08
Q8CGP4	Hist1/2/3	1	2	35.7	35.7	1.55E+08	2.77E+08
Q6A4J8	Usp7	4	27	4.2	27.9	1.22E+07	5.72E+08
Q7TSY8	Sgol2	1	29	1.9	26.6	5.68E+07	5.62E+08
P62806	Hist1h4a	8	12	56.9	61.2	4.81E+07	5.56E+08
Q8CJF7	Ahctf1	3	45	2.3	29.1	1.30E+07	5.37E+08
P20029	Hspa5	28	35	50.5	59.0	1.32E+08	5.28E+08
Q61033	Tmpo	2	27	5.2	46.3	2.91E+07	5.14E+08
Q6NV83	U2surp	31	36	31.2	36.5	2.02E+08	4.91E+08
Q61656	Ddx5	22	23	37.9	38.3	2.19E+08	4.70E+08
Q9EQU5	Set	0	14	0.0	44.9	0.00E+00	4.55E+08
O70133	Dhx9	29	32	23.6	26.3	1.59E+08	4.42E+08
Q99LI5	Znf281	3	20	5.9	34.8	2.64E+07	4.41E+08
Q9ERU9	Ranbp2	25	59	9.5	22.6	7.77E+07	4.33E+08
Q61687	Atrx	1	29	0.8	13.3	2.93E+05	4.25E+08
P13864	Dnmt1	4	38	3.2	24.2	3.15E+07	4.05E+08
Q5SF07	Igf2bp2	8	9	21.6	21.5	3.74E+07	1.01E+08
P43247	Msh2	23	24	27.8	27.5	1.73E+08	3.63E+08
Q69ZR9	Fam208a	3	33	2.4	22.3	2.18E+07	3.62E+08
Q61191	Hcfc1	1	26	0.6	15.6	1.89E+05	3.54E+08
Q8VEK3	Hnrnpu	17	17	24.3	22.6	1.19E+08	3.50E+08
Q09XV5	Chd8	1	34	0.8	20.9	7.57E+05	3.46E+08
Q569Z5	Ddx46	3	35	2.7	32.2	1.15E+06	3.46E+08
G5E870	Trip12	18	37	10.6	23.2	5.98E+07	3.38E+08
Q64010	Crk	1	19	3.3	58.9	6.00E+05	3.32E+08

B)

Protein IDs	Gene names	Unique peptides		Coverage (%)		LFQ Intensity	
		Empty	N-BirA* Uhrf1	Empty	N-BirA* Uhrf1	Empty	N-BirA* Uhrf1
Q8VDF2	Uhrf1	4	56	9.1	55.5	6.27E+06	1.73E+10
Q91ZA3	Pcca	50	55	56.7	62.1	3.31E+09	9.74E+09
A2APB8	Tpx2	4	76	4.6	65.6	3.22E+06	6.05E+09
P63017	Hspa8	41	41	57.9	57.0	2.10E+09	5.36E+09
E9Q4Z2	Acacb	69	91	34.4	40.5	1.15E+09	5.04E+09
P37913	Lig1	13	50	15.7	50.6	2.73E+07	4.90E+09
Q69ZK6	Jmjd1c	29	114	15.8	46.9	6.85E+07	4.71E+09
Q8BX22	Sall4	2	46	1.7	51.0	3.55E+06	4.63E+09
E9PVX6	Mki67	2	164	0.8	54.2	3.48E+05	4.00E+09
Q91ZW3	Smarca5	6	46	8.4	53.4	1.29E+07	3.98E+09
Q99MD9	Nasp	3	45	6.6	55.6	2.64E+06	2.94E+09
Q62318	Trim28	5	40	5.7	40.7	4.13E+06	2.74E+09
Q61937	Npm1	8	25	32.9	60.9	1.61E+07	2.69E+09
Q8BGG6	Mospd3	1	1	8.1	8.1	0.00E+00	2.68E+09
O88491	Nsd1	1	92	0.5	38.6	2.45E+05	2.37E+09
Q8C827	Zfp62	1	33	1.7	35.5	1.29E+07	2.28E+09
Q3U9G9	Lbr	13	34	17.2	35.3	7.17E+07	1.87E+09
Q6PDQ2	Chd4	5	68	3.7	38.6	3.40E+06	1.68E+09
Q6PR54	Rif1	7	75	3.7	32.1	6.71E+06	1.59E+09
P54276	Msh6	16	51	11.9	34.6	2.87E+07	1.52E+09
E9Q7G0	Numa1	22	91	12.6	47.6	2.66E+07	1.40E+09
P09405	Ncl	4	46	8.5	47.5	6.81E+06	1.39E+09
Q6A4J8	Usp7	3	29	3.0	30.9	8.99E+06	1.05E+09
Q80W00	Ppp1r10	0	39	0.0	48.5	0.00E+00	9.84E+08
P59328	Wdhd1	0	35	0.2	38.8	0.00E+00	9.33E+08
Q9EQU5	Set	0	15	0.0	43.3	0.00E+00	9.28E+08
Q61033	Tmpo	1	29	3.1	49.3	1.41E+06	8.81E+08
Q6ZPY7	Kdm3b	6	38	6.0	33.6	2.68E+07	8.65E+08
Q8CGP4	Hist1/2/3	1	2	31.0	35.7	6.48E+07	5.48E+08
Q01320	Top2a	13	49	11.4	35.2	3.21E+07	8.45E+08
P62806	Hist1h4a	6	12	51.8	61.2	2.64E+07	8.20E+08
Q9ESU6	Brd4	0	30	0.0	28.1	0.00E+00	7.97E+08
Q61687	Atrx	2	31	0.8	13.8	9.83E+05	7.54E+08
Q3URK3	Tet1	0	43	0.1	24.0	0.00E+00	7.50E+08
P20029	Hspa5	30	37	52.8	61.0	2.15E+08	7.33E+08
Q569Z5	Ddx46	6	45	5.0	41.1	4.70E+06	7.30E+08
Q8CDN6	Txnl1	5	11	25.3	52.2	1.84E+07	6.79E+08
Q7TSY8	Sgol2	2	27	1.8	25.8	6.27E+07	6.45E+08
Q64010	Crk	1	22	3.6	65.4	0.00E+00	6.42E+08
Q6NV83	U2surp	32	35	32.7	34.5	2.61E+08	6.20E+08
Q99LI5	Znf281	2	25	2.8	42.0	1.02E+06	6.17E+08
Q61656	Ddx5	21	24	33.4	40.8	2.46E+08	6.07E+08
Q9ERU9	Ranbp2	23	71	8.5	25.6	6.98E+07	6.07E+08
P33174	Kif4	1	42	1.4	39.5	4.54E+07	5.98E+08
Q9R1C0	Taf7	1	23	0.9	56.4	9.69E+04	5.97E+08
G5E870	Trip12	20	44	11.8	27.6	7.59E+07	5.94E+08
Q9JLI8	Sart3	1	36	0.7	38.4	5.85E+05	5.94E+08
Q60848	Hells	12	33	15.6	37.5	3.33E+07	5.76E+08
Q65Z40	Wapal	3	34	4.1	32.5	3.16E+06	5.68E+08
Q8CJF7	Ahctf1	4	46	2.2	30.8	1.33E+07	5.67E+08

C)

Protein IDs	Gene names	Unique peptides		Coverage (%)		LFQ Intensity	
		Empty	C-BirA* Uhrf1	Empty	C-BirA* Uhrf1	Empty	C-BirA* Uhrf1
E9Q4Z2	Acacb	80	94	37.1	39.7	1.77E+09	4.82E+09
Q8VDF2	Uhrf1	10	51	20.2	52.2	3.27E+07	4.04E+09
Q69ZK6	Jmjd1c	37	115	19.3	46.7	1.23E+08	3.38E+09
P63017	Hspa8	37	35	60.9	54.8	1.34E+09	2.76E+09
E9PVX6	Mki67	3	155	1.3	50.8	6.32E+05	2.28E+09
E9Q7G0	Numa1	11	115	6.4	53.3	1.23E+07	2.08E+09
Q8C827	Zfp62	1	32	1.7	34.6	1.10E+07	1.86E+09
Q8BX22	Sall4	2	43	2.0	51.8	4.56E+06	1.81E+09
A2APB8	Tpx2	4	69	4.9	64.0	3.09E+06	1.65E+09
Q91ZW3	Smarca5	4	40	7.1	49.5	8.75E+06	1.43E+09
Q6PR54	Rif1	6	73	3.5	30.9	1.10E+07	1.23E+09
O88491	Nsd1	2	83	0.7	36.3	2.72E+06	1.02E+09
P37913	Lig1	9	45	11.7	47.5	1.40E+07	9.05E+08
Q62318	Trim28	5	35	6.8	35.5	3.39E+06	7.55E+08
Q61937	Npm1	5	22	28.1	60.5	4.17E+07	6.91E+08
Q8BGG6	Mospd3	0	1	2.7	8.1	0.00E+00	6.81E+08
Q99MD9	Nasp	3	31	3.0	42.2	9.62E+05	6.62E+08
P09405	Ncl	4	35	9.3	43.7	1.42E+07	5.95E+08
Q6PDQ2	Chd4	3	57	2.7	33.2	4.14E+06	5.48E+08
Q80W00	Ppp1r10	1	32	1.6	39.3	0.00E+00	4.97E+08
Q3U9G9	Lbr	16	26	21.7	31.0	7.60E+07	4.59E+08
P54276	Msh6	13	41	9.3	29.7	1.44E+07	4.40E+08
Q01320	Top2a	13	44	11.5	31.3	3.09E+07	4.11E+08
Q3URK3	Tet1	2	38	0.8	22.7	3.73E+05	3.67E+08
P59328	Wdhd1	0	29	0.4	33.6	0.00E+00	3.03E+08
P62806	Hist1h4a	8	11	56.9	58.0	4.81E+07	2.92E+08
Q9ERU9	Ranbp2	25	56	9.5	21.2	7.77E+07	2.91E+08
Q61033	Tmpo	2	23	5.2	47.7	2.91E+07	2.77E+08
P20029	Hspa5	28	29	50.5	52.1	1.32E+08	2.61E+08
Q7TSY8	Sgol2	1	24	1.9	20.5	5.68E+07	2.59E+08
Q8CGP2	Hist1h/2h	6	9	46.3	53.5	5.41E+07	2.56E+08
Q8CJF7	Ahctf1	3	36	2.3	24.6	1.30E+07	2.45E+08
G5E870	Trip12	18	34	10.6	20.4	5.98E+07	2.32E+08
Q61687	Atrx	1	27	0.8	11.7	2.93E+05	2.10E+08
P13864	Dnmt1	4	35	3.2	21.3	3.15E+07	2.06E+08
Q9EQU5	Set	0	12	0.0	40.5	0.00E+00	2.01E+08
Q99LI5	Znf281	3	21	5.9	35.2	2.64E+07	1.99E+08
Q9ESU6	Brd4	0	22	0.2	19.9	0.00E+00	1.93E+08
P33174	Kif4	1	28	0.6	27.4	0.00E+00	1.66E+08
A2A654	Bptf	1	31	0.4	12.3	0.00E+00	1.59E+08
P84228	Hist1h3a/b/H3f3a/c	3	5	17.2	23.5	3.80E+07	1.50E+08
Q61216	Mre11a	0	17	1.2	30.7	0.00E+00	1.48E+08
Q8BU11	Tox4	0	11	0.0	23.2	0.00E+00	1.47E+08
Q6P8I4	Pcnp	0	9	2.1	38.4	0.00E+00	1.37E+08
Q09XV5	Chd8	1	23	0.8	13.0	7.57E+05	1.35E+08
Q60862	Orc2	0	13	0.0	19.6	0.00E+00	1.32E+08
Q8K298	Anln	0	21	0.0	22.7	0.00E+00	1.22E+08
Q61191	Hcfc1	1	17	0.6	10.5	1.89E+05	1.26E+08
Q6ZPY7	Kdm3b	2	21	1.4	20.6	1.15E+07	1.19E+08
A2AWL7	Mga	0	27	0.0	12.7	0.00E+00	1.19E+08

D)

Protein IDs	Gene names	Unique peptides		Coverage (%)		LFQ Intensity	
		Empty	C-BirA* Uhrf1	Empty	C-BirA* Uhrf1	Empty	C-BirA* Uhrf1
Q8VDF2	Uhrf1	4	55	9.1	55.9	6.27E+06	1.71E+10
Q91ZA3	Pcca	50	57	56.7	63.6	3.31E+09	9.04E+09
Q8BX22	Sall4	2	47	1.7	53.9	3.55E+06	5.35E+09
A2APB8	Tpx2	4	75	4.6	64.6	3.22E+06	5.14E+09
E9Q4Z2	Acacb	69	91	34.4	40.4	1.15E+09	4.26E+09
Q69ZK6	Jmjd1c	29	109	15.8	45.9	6.85E+07	3.39E+09
Q91ZW3	Smarca5	6	43	8.4	51.3	1.29E+07	3.21E+09
Q8C827	Zfp62	1	34	1.7	35.2	1.29E+07	2.63E+09
P37913	Lig1	13	49	15.7	50.1	2.73E+07	2.32E+09
E9PVX6	Mki67	2	157	0.8	51.1	3.48E+05	2.24E+09
E9Q7G0	Numa1	22	101	12.6	50.2	2.66E+07	1.86E+09
Q61937	Npm1	8	24	32.9	59.1	1.61E+07	1.75E+09
O88491	Nsd1	1	91	0.5	38.6	2.45E+05	1.51E+09
Q62318	Trim28	5	39	5.7	40.4	4.13E+06	1.35E+09
Q99MD9	Nasp	3	39	6.6	50.1	2.64E+06	1.29E+09
P09405	Ncl	4	45	8.5	46.5	6.81E+06	1.23E+09
P54276	Msh6	16	47	11.9	33.1	2.87E+07	1.12E+09
Q6PDQ2	Chd4	5	65	3.7	37.0	3.40E+06	1.10E+09
Q80W00	Ppp1r10	0	39	0.0	49.4	0.00E+00	9.79E+08
Q3U9G9	Lbr	13	31	17.2	34.7	7.17E+07	9.71E+08
Q6PR54	Rif1	7	63	3.7	27.5	6.71E+06	9.44E+08
G5E870	Trip12	20	44	11.8	26.6	7.59E+07	6.65E+08
Q61033	Tmpo	1	30	3.1	52.2	1.41E+06	6.41E+08
Q9ERU9	Ranbp2	23	72	8.5	27.3	6.98E+07	6.16E+08
Q9EQU5	Set	0	15	0.0	42.4	0.00E+00	6.06E+08
P62806	Hist1h4a	6	12	51.8	61.2	2.64E+07	5.92E+08
P59328	Wdhd1	0	33	0.2	36.5	0.00E+00	5.91E+08
P20029	Hspa5	30	34	52.8	57.1	2.15E+08	5.81E+08
Q6ZPY7	Kdm3b	6	35	6.0	32.7	2.68E+07	5.77E+08
Q7TSY8	Sgol2	2	30	1.8	27.8	6.27E+07	5.67E+08
Q01320	Top2a	13	45	11.4	32.3	3.21E+07	5.50E+08
Q61687	Atrx	2	28	0.8	12.7	9.83E+05	5.42E+08
Q9ESU6	Brd4	0	26	0.0	23.5	0.00E+00	5.07E+08
Q99LI5	Znf281	2	25	2.8	42.4	1.02E+06	5.06E+08
Q8CGP2	Hist1h/2h	5	10	42.3	59.5	4.22E+07	4.77E+08
P13864	Dnmt1	5	41	3.8	25.4	1.69E+07	4.77E+08
Q9R1C0	Taf7	1	21	0.9	54.7	9.69E+04	4.75E+08
P33174	Kif4	1	41	1.4	37.1	4.54E+07	4.58E+08
Q8CJF7	Ahctf1	4	43	2.2	28.5	1.33E+07	4.51E+08
Q3URK3	Tet1	0	40	0.1	23.3	0.00E+00	4.46E+08
Q8VEK3	Hnrnpu	16	17	21.3	22.3	1.22E+08	4.08E+08
Q569Z5	Ddx46	6	35	5.0	33.2	4.70E+06	3.96E+08
Q60848	Hells	12	30	15.6	35.2	3.33E+07	3.91E+08
Q6P8I4	Pcnp	1	11	7.1	54.7	2.09E+05	3.85E+08
Q8CGP4	Hist1h2/H2	1	2	31.0	35.7	6.48E+07	3.80E+08
Q64010	Crk	1	18	3.6	58.4	0.00E+00	3.53E+08
O88477	Igf2bp1	20	29	39.1	49.3	1.21E+08	3.52E+08
Q8BU11	Tox4	0	14	0.0	29.9	0.00E+00	3.41E+08
Q8K298	Anln	2	31	2.4	32.0	7.88E+05	3.40E+08
Q09XV5	Chd8	2	33	1.0	20.8	2.92E+05	3.22E+08

E)

Protein IDs	Gene names	Unique peptides		Coverage (%)		LFQ Intensity	
		BirA*	N-BirA* Uhrf1	BirA*	N-BirA* Uhrf1	BirA*	N-BirA* Uhrf1
Q5SWU9	Acaca	122	155	55.6	59.6	2.03E+10	8.43E+10
Q99MR8	Mccc1	39	48	55.3	65.1	6.22E+09	1.62E+10
Q91ZA3	Pcca	45	57	54.9	64.5	4.67E+09	1.17E+10
E9Q4Z2	Acacb	30	96	20.2	41.1	7.64E+08	6.80E+09
Q8VDF2	Uhrf1	3	52	4.3	52.8	2.71E+06	6.50E+09
E9PVX6	Mki67	133	168	48.9	54.8	2.10E+09	5.13E+09
O54946	Dnajb6	8	13	21.6	38.4	1.40E+08	2.91E+08
P00405	Mtco2	2	3	10.3	15.3	3.02E+07	1.11E+08
Q6J1H4	Utf1	2	5	7.1	25.5	1.44E+07	8.75E+07
P83887	Tubg1	0	1	9.8	22.6	2.12E+07	5.78E+07
O70445	Bard1	4	10	6.7	14.5	1.85E+07	4.32E+07
Q99J72	Apobec3	0	5	0.7	13.7	0.00E+00	3.94E+07
Q99MN9	Pccb	0	7	0.7	14.0	0.00E+00	3.79E+07
Q99J56	Derl1	1	3	4.7	11.3	7.11E+06	3.76E+07
Q8VE73	Cul7	0	3	0.2	2.3	0.00E+00	3.33E+07
P17012	Zfa/x	1	7	1.7	8.1	5.02E+06	2.95E+07
Q60931	Vdac3	1	4	8.0	14.3	6.37E+06	2.48E+07
Q91W43	Gldc	2	4	2.4	5.4	5.46E+06	2.32E+07
P61089	Ube2n	1	2	7.2	19.7	0.00E+00	2.29E+07
Q8QZY3	Dppa3	0	3	0.0	21.3	0.00E+00	1.82E+07
Q8K4F6	Nsun5	0	4	0.0	9.2	0.00E+00	1.79E+07
Q9D0D4	Dimt1	1	3	3.8	9.1	2.95E+06	1.71E+07
P43276	Hist1h1b	1	3	9.9	19.4	0.00E+00	1.64E+07
Q9D880	Timm50	1	3	4.1	10.4	2.40E+06	1.59E+07
Q8BFV2	Pcid2	1	2	2.7	6.3	0.00E+00	1.57E+07
Q91YI6	Papd4	1	4	2.8	9.7	2.76E+06	1.50E+07
P49615	Cdk5	1	3	4.9	14.6	0.00E+00	1.42E+07
Q8JZM0	Tfb1m	1	2	3.2	9.6	0.00E+00	1.39E+07
O88291	Znf326	1	3	1.9	6.4	0.00E+00	1.31E+07
Q8BFP9	Pdk1	0	3	1.2	10.6	0.00E+00	1.28E+07
Q9DC33	Hmg20a	1	3	5.2	10.8	0.00E+00	1.25E+07
Q9QXX4	Slc25a13	1	3	1.6	4.8	0.00E+00	1.03E+07
O88986	Gcat	0	2	1.0	7.9	0.00E+00	9.47E+06
Q9CQS5	Riok2	0	3	0.7	6.5	0.00E+00	8.73E+06
Q99L48	Nmd3	0	2	0.0	6.2	0.00E+00	8.50E+06
Q67FY2	Bcl9l	1	2	0.7	3.1	0.00E+00	7.26E+06
Q9QZD4	Ercc4	1	2	2.6	2.7	1.40E+06	6.46E+06
Q8JZU2	Slc25a1	0	3	1.2	9.8	0.00E+00	6.05E+06
Q8R0W0	Eppk1	0	2	0.0	3.9	0.00E+00	5.52E+06
Q91VA7	Idh3b	1	2	2.2	5.3	0.00E+00	4.71E+06
Q9CWX9	Ddx47	1	3	2.6	6.9	0.00E+00	4.71E+06
Q9ES63	Usp29	0	2	0.8	4.4	0.00E+00	4.10E+06
Q64G17	Anp32a/b/c	0	2	2.2	13.0	0.00E+00	3.86E+06

F)

Protein IDs	Gene names	Unique peptides		Coverage (%)		LFQ Intensity	
		BirA*	N-BirA* Uhrf1	BirA*	N-BirA* Uhrf1	BirA*	N-BirA* Uhrf1
Q5SWU9	Acaca	118	153	54.2	59.3	1.67E+10	5.52E+10
Q8VDF2	Uhrf1	3	56	4.4	55.5	4.74E+06	1.73E+10
Q99MR8	Mccc1	38	48	54.6	66.5	5.42E+09	1.48E+10
Q91ZA3	Pcca	40	55	52.8	62.1	2.98E+09	9.74E+09
E9Q4Z2	Acacb	28	91	19.3	40.5	5.69E+08	5.04E+09
E9PVX6	Mki67	104	164	39.9	54.2	1.05E+09	4.00E+09
E9Q7G0	Numa1	44	91	26.8	47.6	7.02E+08	1.40E+09
Q3URK3	Tet1	24	43	15.1	24.0	2.18E+08	7.50E+08
Q8CDN6	Txnl1	9	11	38.2	52.2	3.44E+08	6.79E+08
O54946	Dnajb6	8	11	22.3	31.8	1.41E+08	3.24E+08
Q8QZY3	Dppa3	0	6	2.0	34.0	0.00E+00	1.91E+08
Q80TZ9	Rere	5	22	4.0	17.9	3.25E+07	1.88E+08
Q9EPW0	Inpp4a	1	1	2.1	2.1	2.38E+07	1.15E+08
O70445	Bard1	5	12	8.4	18.4	3.55E+07	1.08E+08
Q8BRB7	Kat6b	3	11	2.3	8.5	2.84E+07	6.62E+07
A0A087WPF7	Auts2	3	12	2.6	10.6	5.20E+06	5.26E+07
Q6P5D8	Smchd1	3	15	2.1	9.3	1.06E+07	5.25E+07
O55028	Bckdk	3	5	8.7	14.8	1.51E+07	3.90E+07
Q6J1H4	Utf1	1	5	1.6	23.5	0.00E+00	3.81E+07
P17012	Zfa/x	0	7	0.4	8.4	0.00E+00	3.69E+07
A2AUY4	Baz2b	1	8	1.0	7.3	6.10E+06	3.63E+07
Q99MN9	Pccb	0	5	0.7	10.0	0.00E+00	3.29E+07
P61089	Ube2n	1	2	11.4	19.7	6.42E+06	2.50E+07
P68510	Ywhah	1	2	11.2	21.7	0.00E+00	2.37E+07
O88286	Wiz	1	7	1.1	5.7	5.06E+06	2.14E+07
Q920L1	Fads1	0	3	0.0	8.7	0.00E+00	1.95E+07
A0A087WP63	Srcap	1	5	0.3	2.3	0.00E+00	1.72E+07
Q8K4F6	Nsun5	1	3	1.5	7.6	0.00E+00	1.68E+07
Q8BFV2	Pcid2	1	2	4.0	6.3	0.00E+00	1.46E+07
Q64G17	Anp32a/b/c	1	2	8.7	13.0	0.00E+00	1.37E+07
Q91VW9	Zkscan3	1	2	1.2	4.6	0.00E+00	1.36E+07
Q61550	Rad21	0	4	0.0	7.1	0.00E+00	1.13E+07
Q9QZD4	Ercc4	0	3	0.4	4.0	0.00E+00	1.07E+07
Q9CQS5	Riok2	1	3	2.1	5.8	0.00E+00	9.43E+06
P19324	Serpinh1	0	2	0.0	7.0	0.00E+00	8.82E+06
P0CG14	Chtf8	1	3	0.9	12.4	0.00E+00	8.74E+06
Q6ZPV2	Ino80	0	2	0.0	1.6	0.00E+00	8.69E+06
Q3ULD5	Mccc2	1	4	2.9	8.2	6.89E+05	8.52E+06
Q924W5	Smc6	1	3	0.8	3.8	0.00E+00	8.40E+06
P58283	Rnf216	1	3	2.3	4.9	0.00E+00	7.53E+06
Q99L48	Nmd3	1	2	3.2	6.9	0.00E+00	6.47E+06
O70503	Hsd17b12	0	2	1.0	8.0	0.00E+00	6.21E+06
Q99M31	Hspa14	0	2	1.4	7.6	0.00E+00	6.09E+06
Q91VA7	Idh3b	1	3	3.1	5.3	0.00E+00	5.27E+06
O88508	Dnmt3a	0	1	2.0	3.4	0.00E+00	4.87E+06

G)

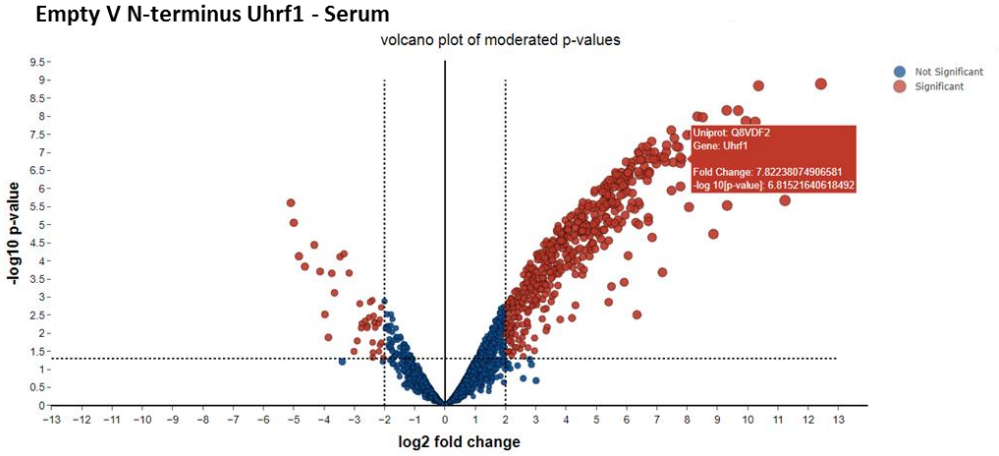
Protein IDs	Gene names	Unique peptides		Coverage (%)		LFQ Intensity	
		BirA*	C-BirA* Uhrf1	BirA*	C-BirA* Uhrf1	BirA*	C-BirA* Uhrf1
Q5SWU9	Acaca	122	154	55.6	58.5	2.03E+10	5.53E+10
Q99MR8	Mccc1	39	48	55.3	65.7	6.22E+09	1.23E+10
E9Q4Z2	Acacb	30	94	20.2	39.7	7.64E+08	4.82E+09
Q8VDF2	Uhrf1	3	51	4.3	52.2	2.71E+06	4.04E+09
Q80TZ9	Rere	5	20	5.6	16.5	3.95E+07	1.08E+08
Q6J1H4	Utf1	2	6	7.1	26.8	1.44E+07	8.57E+07
Q99MN9	Pccb	0	6	0.7	11.9	0.00E+00	3.32E+07
Q8VE73	Cul7	0	3	0.2	1.8	0.00E+00	2.57E+07
Q99J72	Apobec3	0	4	0.7	11.5	0.00E+00	2.50E+07
Q8QZY3	Dppa3	0	3	0.0	23.1	0.00E+00	1.87E+07
Q3ULD5	Mccc2	1	5	1.5	13.0	0.00E+00	1.70E+07
Q91YI6	Papd4	1	4	2.8	9.7	2.76E+06	1.22E+07
P43276	Hist1h1b	1	3	9.9	19.7	0.00E+00	1.12E+07
Q8K4F6	Nsun5	0	4	0.0	11.7	0.00E+00	1.00E+07
Q9D880	Timm50	1	4	4.1	10.4	2.40E+06	9.90E+06
Q1H9T6	Zbtb48	1	2	1.0	3.1	0.00E+00	8.92E+06
Q8BFP9	Pdk1	0	3	1.2	9.4	0.00E+00	8.39E+06
Q9DC33	Hmg20a	1	2	5.2	9.8	0.00E+00	8.38E+06
P49615	Cdk5	1	3	4.9	13.4	0.00E+00	7.77E+06
P59235	Nup43	1	2	2.7	6.8	0.00E+00	7.60E+06
Q9QXX4	Slc25a13	1	4	1.6	5.7	0.00E+00	7.10E+06
Q9CQS5	Riok2	0	2	0.7	4.6	0.00E+00	7.07E+06
Q99L48	Nmd3	0	2	0.0	6.2	0.00E+00	5.66E+06
Q67FY2	Bcl9l	1	3	0.7	2.8	0.00E+00	4.98E+06
Q61602	Gli3	1	2	2.1	3.2	0.00E+00	4.87E+06
Q8JZU2	Slc25a1	0	2	1.2	7.4	0.00E+00	4.83E+06
Q6PGG6	Gnl3l	2	0	6.6	0.0	8.92E+06	0.00E+00
P17751	Tpi1	0	2	0.0	8.4	0.00E+00	1.88E+06

H)

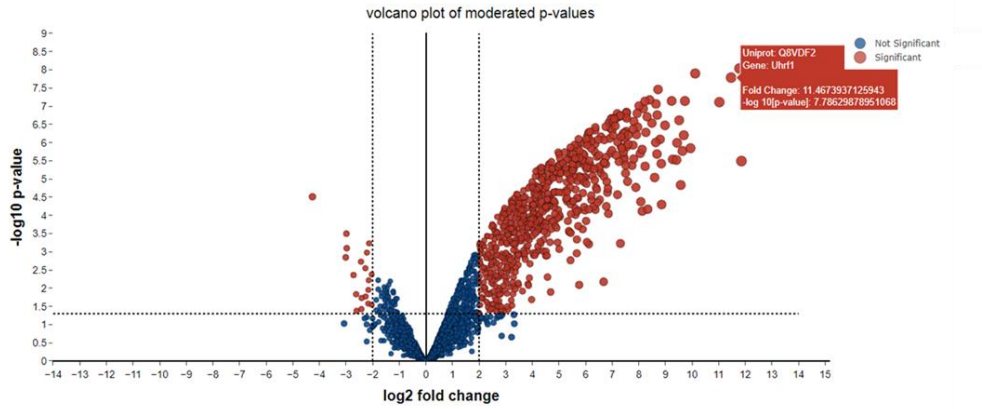
Protein IDs	Gene names	Unique peptides		Coverage (%)		LFQ Intensity	
		BirA*	C-BirA* Uhrf1	BirA*	C-BirA* Uhrf1	BirA*	C-BirA* Uhrf1
Q8VDF2	Uhrf1	3	55	4.4	55.9	4.74E+06	1.71E+10
Q99MR8	Mccc1	38	50	54.6	66.4	5.42E+09	1.38E+10
Q91ZA3	Pcca	40	57	52.8	63.6	2.98E+09	9.04E+09
E9Q4Z2	Acacb	28	91	19.3	40.4	5.69E+08	4.26E+09
E9PVX6	Mki67	104	157	39.9	51.1	1.05E+09	2.24E+09
E9Q7G0	Numa1	44	101	26.8	50.2	7.02E+08	1.86E+09
O54946	Dnajb6	8	12	22.3	35.1	1.41E+08	3.58E+08
Q8QZY3	Dppa3	0	8	2	36.7	0.00E+00	1.95E+08
Q80TZ9	Rere	5	23	4	18.4	3.25E+07	1.79E+08
A0A087WPF7	Auts2	3	13	2.6	12	5.20E+06	5.38E+07
Q6J1H4	Utf1	1	5	1.6	20.4	0.00E+00	4.29E+07
Q99MN9	Pccb	0	7	0.7	12.2	0.00E+00	3.48E+07
A2AUY4	Baz2b	1	9	1	7.5	6.10E+06	3.40E+07
O88286	Wiz	1	7	1.1	5.7	5.06E+06	2.43E+07
Q920L1	Fads1	0	2	0	7.2	0.00E+00	1.70E+07
Q9QYI5	Dnajb2	1	2	2.5	10.9	0.00E+00	1.43E+07
Q8K4F6	Nsun5	1	4	1.5	11.5	0.00E+00	1.36E+07
P68510	Ywhah	1	2	11.2	19.4	0.00E+00	1.34E+07
E9Q9T0	Fbrsl1	0	3	0	7.1	0.00E+00	1.29E+07
P17012	Zfa/x	0	6	0.4	7.4	0.00E+00	1.11E+07
A0A087WP63	Srcap	1	5	0.3	2.2	0.00E+00	1.03E+07
P0CG14	Chtf8	1	4	0.9	14.1	0.00E+00	1.03E+07
Q7TMY8	Huwe1	1	3	0.3	1.2	0.00E+00	1.02E+07
Q9JMK2	Csnk1e	1	3	3.1	8.3	0.00E+00	9.16E+06
Q9CQS5	Riok2	1	3	2.1	5.8	0.00E+00	8.76E+06
B8JJZ9	Zfp563	0	2	1.3	6.2	0.00E+00	8.70E+06
Q61550	Rad21	0	3	0	6.6	0.00E+00	8.25E+06
P59235	Nup43	0	2	1.5	7	0.00E+00	7.97E+06
O88986	Gcat	0	3	1	8.6	0.00E+00	6.56E+06
P58283	Rnf216	1	3	2.3	4.5	0.00E+00	6.14E+06
Q99L48	Nmd3	1	2	3.2	7.1	0.00E+00	5.87E+06
P19324	Serpinh1	0	2	0	8.2	0.00E+00	5.67E+06
Q9CR62	Slc25a11	0	2	0	5.4	0.00E+00	5.35E+06
G3X972	Sec24c	0	2	0.4	2.4	0.00E+00	5.00E+06
Q64G17	Anp32a/b/c	1	2	8.7	13	0.00E+00	4.77E+06

Appendix 2-Volcano plots of proteins enriched in the Uhrf1 BioID screen relative to control lines

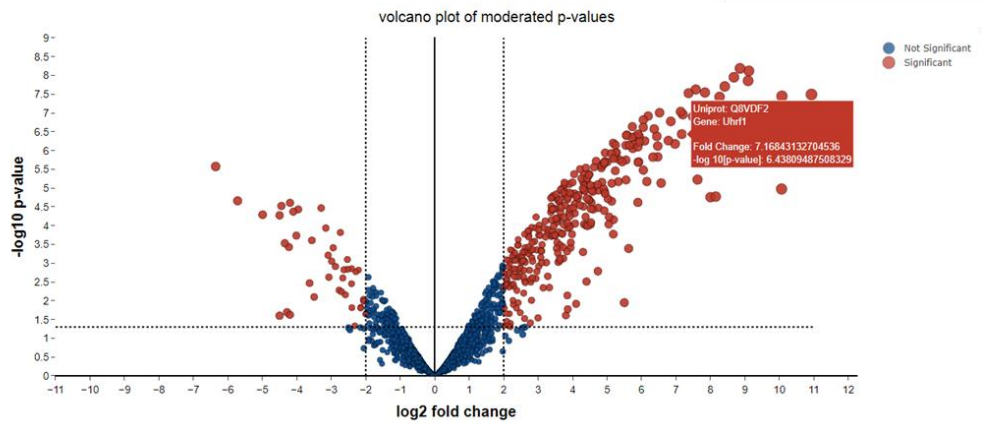
Empty V N-terminus Uhrf1 - Serum



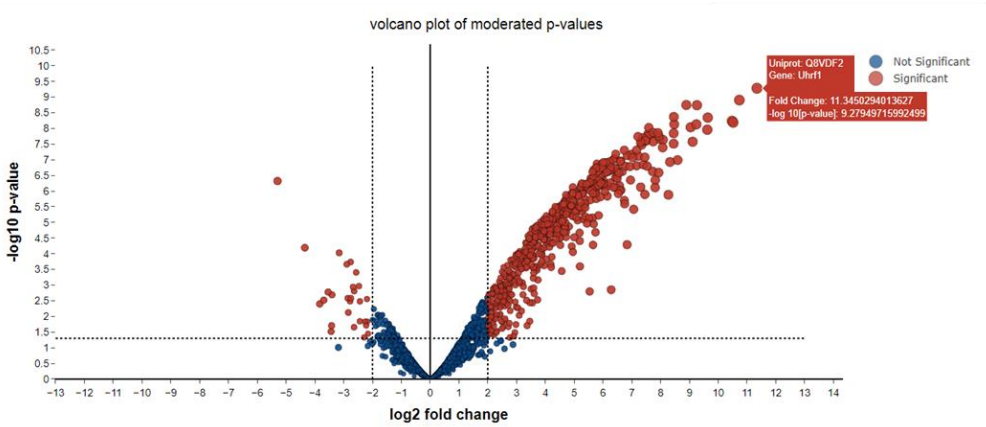
Empty V N-terminus Uhrf1 – KOSR+2i



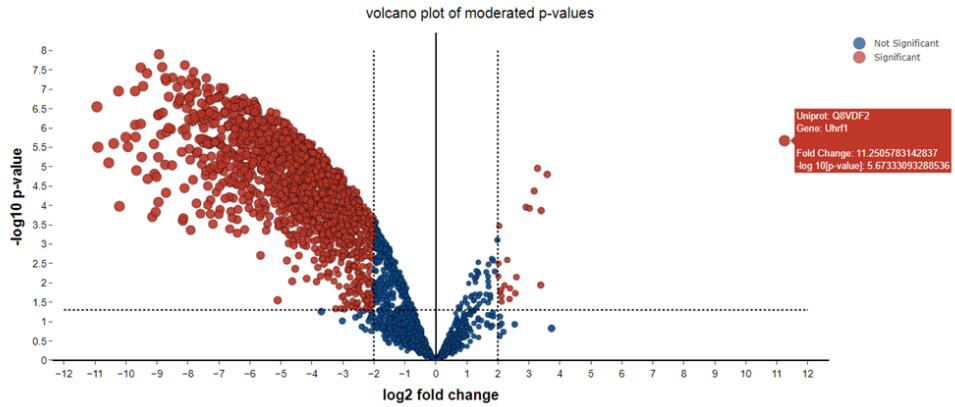
Empty V C-terminus Uhrf1 – Serum



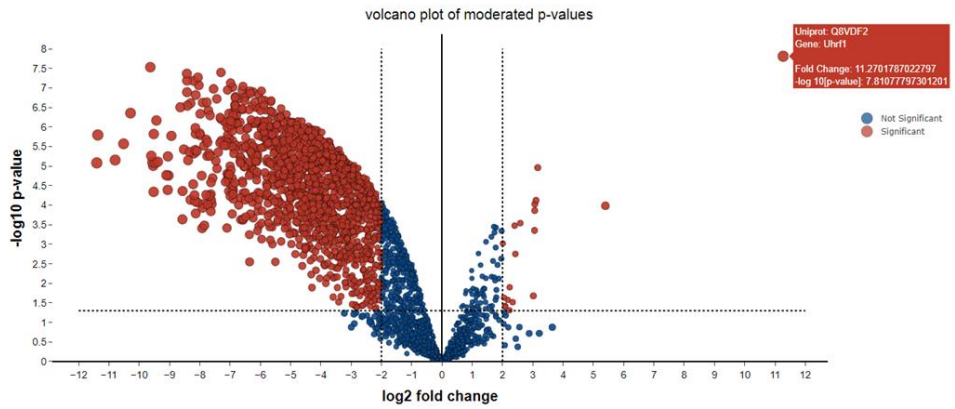
Empty V C-terminus Uhrf1 – KOSR+2i



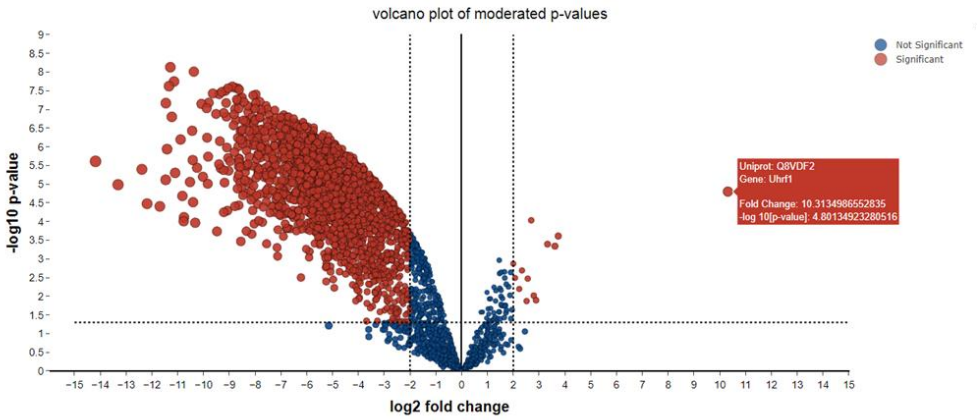
BirA* V N-terminus Uhrf1 – Serum



BirA* V N-terminus Uhrf1 – KOSR+2i



BirA* V C-terminus Uhrf1 – Serum



BirA* V C-terminus Uhrf1 – KOSR+2i

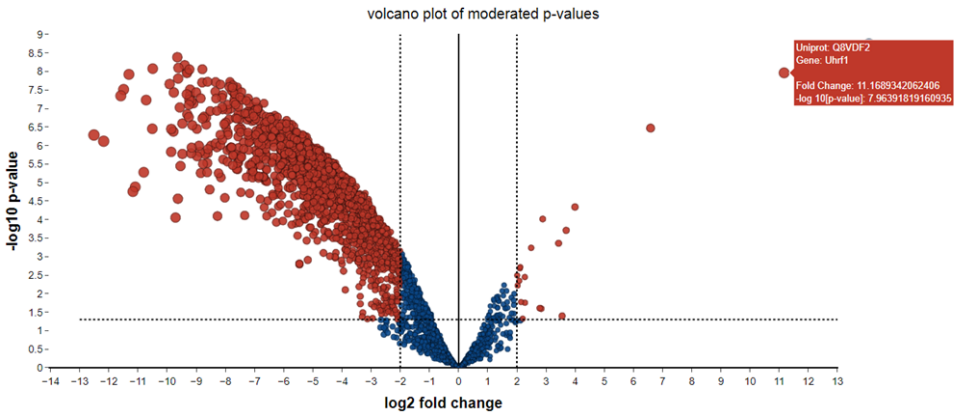


Figure Appendix 2 – Volcano plots illustrating the relative enrichment of each protein detected in the Uhrf1 BioID screen for the conditions labelled. Missing values for each protein were imputed and *p* values were calculated using Limma moderated statistics. Plots were generated by Jimi Wills (IGMM, University of Edinburgh)

Bibliography

Alabert, C. *et al.* (2014) 'Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components', *Nature Cell Biology*, 16(3), pp. 281–291. doi: 10.1038/ncb2918.

An, L. *et al.* (2018) 'RNF169 limits 53BP1 deposition at DSBs to stimulate single-strand annealing repair', *Proceedings of the National Academy of Sciences*, 115(35), pp. E8286–E8295. doi: 10.1073/pnas.1804823115.

Arand, J. *et al.* (2012) 'In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases', *PLoS Genetics*. Edited by D. Schübeler, 8(6), p. e1002750. doi: 10.1371/journal.pgen.1002750.

Arand, J. *et al.* (2015) 'Selective impairment of methylation maintenance is the major cause of DNA methylation reprogramming in the early embryo', *Epigenetics & Chromatin*, 8(1), p. 1. doi: 10.1186/1756-8935-8-1.

Arita, K. *et al.* (2008) 'Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism', *Nature*, 455(7214), pp. 818–821. doi: 10.1038/nature07249.

Avvakumov, G. V. *et al.* (2008) 'Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1', *Nature*, 455(7214), pp. 822–825. doi: 10.1038/nature07273.

Bao, Q. *et al.* (2017) 'Utf1 contributes to intergenerational epigenetic inheritance of pluripotency', *Scientific Reports*. Nature Publishing Group, 7(1), p. 14612. doi: 10.1038/s41598-017-14426-5.

Barau, J. *et al.* (2016) 'The DNA methyltransferase DNMT3C protects male germ cells from transposon activity', *Science*, 354(6314), pp. 909–912. doi: 10.1126/science.aah5143.

Barr, H. *et al.* (2007) 'Mbd2 Contributes to DNA Methylation-Directed Repression of the Xist Gene', *Molecular and Cellular Biology*, 27(10), pp. 3750–3757. doi: 10.1128/MCB.02204-06.

Bashtrykov, P. *et al.* (2012) 'Specificity of Dnmt1 for Methylation of Hemimethylated CpG Sites Resides in Its Catalytic Domain', *Chemistry & Biology*, 19(5), pp. 572–578. doi: 10.1016/j.chembiol.2012.03.010.

Bashtrykov, P., Rajavelu, A., *et al.* (2014) 'Targeted Mutagenesis Results in an Activation of DNA Methyltransferase 1 and Confirms an Autoinhibitory Role of its RFTS Domain', *ChemBioChem*, 15(5), pp. 743–748. doi: 10.1002/cbic.201300740.

Bashtrykov, P., Jankevicius, G., *et al.* (2014) 'The UHRF1 Protein Stimulates the Activity and Specificity of the Maintenance DNA Methyltransferase DNMT1 by an Allosteric Mechanism', *Journal of Biological Chemistry*, 289(7), pp. 4106–4115. doi: 10.1074/jbc.M113.528893.

Baubec, T. *et al.* (2015) 'Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation', *Nature*, 520(7546), pp. 243–247. doi: 10.1038/nature14176.

Baubec, T. and Schübeler, D. (2014) 'Genomic patterns and context specific interpretation of DNA methylation', *Current Opinion in Genetics & Development*, 25, pp. 85–92. doi: 10.1016/j.gde.2013.11.015.

Becker, P. B. and Workman, J. L. (2013) 'Nucleosome Remodeling and Epigenetics', *Cold Spring Harbor Perspectives in Biology*, 5(9), pp. a017905–a017905. doi: 10.1101/cshperspect.a017905.

Behal, R. H. *et al.* (1993) 'Regulation of the Pyruvate Dehydrogenase Multienzyme Complex', *Annual Review of Nutrition*, 13(1), pp. 497–520. doi: 10.1146/annurev.nu.13.070193.002433.

Beraldi, R. *et al.* (2006) 'Expression of LINE-1 retroposons is essential for murine preimplantation development', *Molecular Reproduction and Development*, 73(3), pp. 279–287. doi: 10.1002/mrd.20423.

Bestor, T. H. (2003) 'Unanswered Questions about the Role of Promoter Methylation in Carcinogenesis', *Annals of the New York Academy of Sciences*, 983(1), pp. 22–27. doi: 10.1111/j.1749-6632.2003.tb05959.x.

Bewick, A. J. *et al.* (2016) 'Evolution of DNA Methylation across Insects', *Molecular Biology and Evolution*, 34(3), p. msw264. doi: 10.1093/molbev/msw264.

Bird, A. (2002) 'DNA methylation patterns and epigenetic memory', *Genes & Development*, 16(1), pp. 6–21. doi: 10.1101/gad.947102.

Blaschke, K. *et al.* (2013) 'Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells.', *Nature*. NIH Public Access, 500(7461), pp. 222–6. doi: 10.1038/nature12362.

van den Boom, V. *et al.* (2007) 'UTF1 is a chromatin-associated protein involved in ES cell differentiation', *The Journal of Cell Biology*, 178(6), pp. 913–924. doi: 10.1083/jcb.200702058.

Borgel, J. *et al.* (2010) 'Targets and dynamics of promoter DNA methylation during early mouse development', *Nature Genetics*, 42(12), pp. 1093–1100. doi: 10.1038/ng.708.

Bostick, M. *et al.* (2007) 'UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells', *Science*, 317(5845), pp. 1760–1764. doi: 10.1126/science.1147939.

Bourc'his, D. *et al.* (2001) 'Dnmt3L and the Establishment of Maternal Genomic Imprints', *Science*, 294(5551), pp. 2536–2539. doi: 10.1126/science.1065848.

Bourc'his, D. and Bestor, T. H. (2004) 'Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L', *Nature*, 431(7004), pp. 96–99. doi: 10.1038/nature02886.

Bowker-Kinley, M. and Popov, K. M. (1999) 'Evidence that pyruvate dehydrogenase kinase belongs to the ATPase/kinase superfamily.', *The Biochemical journal*, 344 Pt 1, pp. 47–53. doi: 10.1042/BJ3440047.

Brahmajosyula, M. and Miyake, M. (2013) 'Localization and expression of peptidylarginine deiminase 4 (PAD4) in mammalian oocytes and preimplantation embryos', *Zygote*, 21(4), pp. 314–324. doi: 10.1017/S0967199411000633.

Brandeis, M. *et al.* (1994) 'Spl elements protect a CpG island from de novo methylation', *Nature*, 371(6496), pp. 435–438. doi: 10.1038/371435a0.

Branon, T. C. *et al.* (2018a) 'Efficient proximity labeling in living cells and organisms with TurboID', *Nature Biotechnology*. Nature Publishing Group, 36(9), pp. 880–887. doi: 10.1038/nbt.4201.

Branon, T. C. *et al.* (2018b) 'Efficient proximity labeling in living cells and organisms with TurboID', *Nature Biotechnology*, 36(9), pp. 880–887. doi: 10.1038/nbt.4201.

Brons, I. G. M. *et al.* (2007) 'Derivation of pluripotent epiblast stem cells from mammalian embryos', *Nature*, 448(7150), pp. 191–195. doi: 10.1038/nature05950.

Buehr, M. *et al.* (2008) 'Capture of Authentic Embryonic Stem Cells from Rat Blastocysts', *Cell*, 135(7), pp. 1287–1298. doi: 10.1016/j.cell.2008.12.007.

Bulut-Karslioglu, A. *et al.* (2014) 'Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells', *Molecular Cell*, 55(2), pp. 277–290. doi: 10.1016/j.molcel.2014.05.029.

Catania, S. *et al.* (2017) 'Epigenetic maintenance of DNA methylation after evolutionary loss of the de novo methyltransferase', *bioRxiv*. Cold Spring Harbor Laboratory, p. 149385. doi: 10.1101/149385.

Chahil, G., Yelam, A. and Bollu, P. C. (2018) 'Rett Syndrome in Males: A Case Report and Review of Literature.', *Cureus*. Cureus Inc., 10(10), p. e3414. doi: 10.7759/cureus.3414.

Chang, X. *et al.* (2009) 'Increased PADI4 expression in blood and tissues of patients with malignant tumors', *BMC Cancer*, 9(1), p. 40. doi: 10.1186/1471-2407-9-40.

Charlton, J. *et al.* (2018a) 'Global delay in nascent strand DNA methylation', *Nature Structural & Molecular Biology*. Nature Publishing Group, 25(4), pp. 327–332. doi: 10.1038/s41594-018-0046-4.

Charlton, J. *et al.* (2018b) 'Global delay in nascent strand DNA methylation', *Nature Structural & Molecular Biology*. Nature Publishing Group, 25(4), pp. 327–332. doi: 10.1038/s41594-018-0046-4.

Chaudhry, M. A. *et al.* (2008) 'Basal medium composition and serum or serum replacement concentration influences on the maintenance of murine embryonic stem cells', *Cytotechnology*, 58(3), pp. 173–179. doi: 10.1007/s10616-008-9177-5.

Chen, H. *et al.* (2013) 'DNA Damage Regulates UHRF1 Stability via the SCF $^{\beta\text{-TrCP}}$ E3 Ligase', *Molecular and Cellular Biology*, 33(6), pp. 1139–1148. doi: 10.1128/MCB.01191-12.

Chen, R. Z. *et al.* (2001) 'Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice', *Nature Genetics*, 27(3), pp. 327–331. doi: 10.1038/85906.

Chen, T. *et al.* (2003) 'Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b.', *Molecular and cellular biology*, 23(16), pp. 5594–605. doi: 10.1128/mcb.23.16.5594-5605.2003.

Chiappinelli, K. B. *et al.* (2015) 'Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses', *Cell*, 162(5), pp. 974–986. doi: 10.1016/j.cell.2015.07.011.

Choi, J. *et al.* (2017) 'Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells', *Nature*, 548(7666), pp. 219–223. doi: 10.1038/nature23274.

Christophorou, M. A. *et al.* (2014) 'Citrullination regulates pluripotency and histone H1 binding to chromatin.', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 507(7490), pp. 104–8. doi: 10.1038/nature12942.

Chu, J. *et al.* (2012) 'UHRF1 phosphorylation by cyclin A2/cyclin-dependent kinase 2 is required for zebrafish embryogenesis.', *Molecular biology of the cell*, 23(1), pp. 59–70. doi: 10.1091/mbc.E11-06-0487.

Chuykin, I. A. *et al.* (2008) 'Activation of DNA damage response signaling in mouse embryonic stem cells', *Cell Cycle*, 7(18), pp. 2922–2928. doi: 10.4161/cc.7.18.6699.

Cirio, M. C. *et al.* (2008) 'Preimplantation expression of the somatic form of Dnmt1 suggests a role in the inheritance of genomic imprints', *BMC Developmental Biology*, 8(1), p. 9. doi: 10.1186/1471-213X-8-9.

Clancy, K. W., Weerapana, E. and Thompson, P. R. (2016) 'Detection and identification of protein citrullination in complex biological systems.', *Current opinion in chemical biology*. NIH Public Access, 30, pp. 1–6. doi: 10.1016/j.cbpa.2015.10.014.

Conlan, B. *et al.* (2018) 'Development of a Rapid in planta BioID System as a Probe for Plasma Membrane-Associated Immunity Proteins', *Frontiers in Plant Science*, 9, p. 1882. doi: 10.3389/fpls.2018.01882.

Cortázar, D. *et al.* (2011) 'Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability', *Nature*, 470(7334), pp. 419–423. doi: 10.1038/nature09672.

Cortellino, S. *et al.* (2011) 'Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair', *Cell*, 146(1), pp. 67–79. doi: 10.1016/j.cell.2011.06.020.

Cox, J. *et al.* (2014) 'Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ', *Molecular & Cellular Proteomics*, 13(9), pp. 2513–2526. doi: 10.1074/mcp.M113.031591.

Cox, J. and Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology*, 26(12), pp. 1367–1372. doi: 10.1038/nbt.1511.

Crichton, J. H. *et al.* (2014) 'Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline', *Cellular and Molecular Life Sciences*, 71(9), pp. 1581–1605. doi: 10.1007/s00018-013-1468-0.

Cruickshanks, H. A. *et al.* (2013) 'Senescent cells harbour features of the cancer epigenome', *Nature Cell Biology*, 15(12), pp. 1495–1506. doi: 10.1038/ncb2879.

Dan, J. *et al.* (2017) 'Zscan4 Inhibits Maintenance DNA Methylation to Facilitate Telomere Elongation in Mouse Embryonic Stem Cells', *Cell Reports*, 20(8), pp. 1936–1949. doi: 10.1016/j.celrep.2017.07.070.

de Greef, J. C. *et al.* (2011) 'Mutations in ZBTB24 Are Associated with Immunodeficiency, Centromeric Instability, and Facial Anomalies Syndrome Type 2', *The American Journal of Human Genetics*, 88(6), pp. 796–804. doi: 10.1016/j.ajhg.2011.04.018.

Deplus, R. *et al.* (2014) 'Citrullination of DNMT3A by PADI4 regulates its stability and controls DNA methylation', *Nucleic Acids Research*, 42(13), pp. 8285–8296. doi: 10.1093/nar/gku522.

Dillon, S. C. *et al.* (2005) 'The SET-domain protein superfamily: protein lysine methyltransferases.', *Genome Biology*, 6(8), p. 227. doi: 10.1186/gb-2005-6-8-227.

Domcke, S. *et al.* (2015) 'Competition between DNA methylation and transcription factors determines binding of NRF1', *Nature*, 528(7583), pp. 575–579. doi: 10.1038/nature16462.

Dong, K. B. *et al.* (2008) 'DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity', *The EMBO Journal*, 27(20), pp. 2691–2701. doi: 10.1038/emboj.2008.193.

Du, J., Johnson, Lianna M., *et al.* (2015) 'DNA methylation pathways and their crosstalk with histone methylation', *Nature Reviews Molecular Cell Biology*, 16(9), pp. 519–532. doi: 10.1038/nrm4043.

Du, J., Johnson, Lianna M, *et al.* (2015) 'DNA methylation pathways and their crosstalk with histone methylation', *Nature Publishing Group*, 16. doi: 10.1038/nrm4043.

Du, W. *et al.* (2019) 'Stella protein facilitates DNA demethylation by disrupting the chromatin association of the RING finger-type E3 ubiquitin ligase UHRF1.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 294(22), pp. 8907–8917. doi: 10.1074/jbc.RA119.008008.

Du, Z. *et al.* (2010) 'DNMT1 Stability Is Regulated by Proteins Coordinating Deubiquitination and Acetylation-Driven Ubiquitination', *Science Signaling*, 3(146), pp. ra80–ra80. doi: 10.1126/scisignal.2001462.

Dunican, D. S. *et al.* (2013) 'Lsh regulates LTR retrotransposon repression independently of Dnmt3b function', *Genome Biology*. BioMed Central, 14(12), p. R146. doi: 10.1186/gb-2013-14-12-r146.

Dunican, D. S., Pennings, S. and Meehan, R. R. (2015) 'Lsh Is Essential for Maintaining Global DNA Methylation Levels in Amphibia and Fish and Interacts Directly with Dnmt1', *BioMed Research International*, 2015, pp. 1–12. doi: 10.1155/2015/740637.

Eden, A. *et al.* (2003) 'Chromosomal Instability and Tumors Promoted by DNA Hypomethylation', *Science*, 300(5618), pp. 455–455. doi: 10.1126/science.1083557.

Ehrlich, M. (2002) 'DNA methylation in cancer: too much, but also too little', *Oncogene*, 21(35), pp. 5400–5413. doi: 10.1038/sj.onc.1205651.

Epsztejn-Litman, S. *et al.* (2008) 'De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes', *Nature Structural & Molecular Biology*, 15(11), pp. 1176–1183. doi: 10.1038/nsmb.1476.

Esnault, C. *et al.* (2005) 'APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses', *Nature*, 433(7024), pp. 430–433. doi: 10.1038/nature03238.

Esteve, P.-O. *et al.* (2006) 'Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication', *Genes & Development*, 20(22), pp. 3089–3103. doi: 10.1101/gad.1463706.

Esteve, P.-O. *et al.* (2009) 'Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells', *Proceedings of the National Academy of Sciences*, 106(13), pp. 5076–5081. doi: 10.1073/pnas.0810362106.

Estève, P.-O. *et al.* (2011) 'A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability', *Nature Structural & Molecular Biology*, 18(1), pp. 42–48. doi: 10.1038/nsmb.1939.

Fang, J. *et al.* (2016) 'Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition', *Nature Communications*, 7(1), p. 11197. doi: 10.1038/ncomms11197.

Fatemi, M. *et al.* (2001) 'The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA', *Journal of Molecular Biology*, 309(5), pp. 1189–1199. doi: 10.1006/jmbi.2001.4709.

Fazzio, T. G., Huff, J. T. and Panning, B. (2008) 'An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity', *Cell*, 134(1), pp. 162–174. doi: 10.1016/j.cell.2008.05.031.

Feldman, N. *et al.* (2006) 'G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis', *Nature Cell Biology*, 8(2), pp. 188–194. doi: 10.1038/ncb1353.

Feldmann, A. *et al.* (2013) 'Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions', *PLoS Genetics*. Edited by G. S. Barsh. Public Library of Science, 9(12), p. e1003994. doi: 10.1371/journal.pgen.1003994.

Felle, M. *et al.* (2011) 'The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1', *Nucleic Acids Research*, 39(19), pp. 8355–8365. doi: 10.1093/nar/gkr528.

Fellinger, K. *et al.* (2009) 'Dimerization of DNA methyltransferase 1 is mediated by its regulatory domain', *Journal of Cellular Biochemistry*, 106(4), pp. 521–528. doi: 10.1002/jcb.22071.

Ferguson-Smith, A. C. (2011) 'Genomic imprinting: the emergence of an epigenetic paradigm', *Nature Reviews Genetics*, 12(8), pp. 565–575. doi: 10.1038/nrg3032.

Ferrón, S. R. *et al.* (2011) 'Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis', *Nature*, 475(7356), pp. 381–385. doi: 10.1038/nature10229.

Ferry, L. *et al.* (2017) 'Methylation of DNA Ligase 1 by G9a/GLP Recruits UHRF1 to Replicating DNA and Regulates DNA Methylation', *Molecular Cell*, 67(4), pp. 550-565.e5. doi: 10.1016/j.molcel.2017.07.012.

Ficz, G. *et al.* (2013) 'FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency.', *Cell stem cell*. Elsevier, 13(3), pp. 351–9. doi: 10.1016/j.stem.2013.06.004.

Foster, B. M. *et al.* (2018) 'Critical Role of the UBL Domain in Stimulating the E3 Ubiquitin Ligase Activity of UHRF1 toward Chromatin', *Molecular Cell*. Cell Press, 72(4), pp. 739-752.e9. doi: 10.1016/J.MOLCEL.2018.09.028.

Fournier, A. *et al.* (2012) 'The role of methyl-binding proteins in chromatin organization and epigenome maintenance', *Briefings in Functional Genomics*, 11(3), pp. 251–264. doi: 10.1093/bfgp/elr040.

Fouse, S. D. *et al.* (2008) 'Promoter CpG Methylation Contributes to ES Cell Gene Regulation in Parallel with Oct4/Nanog, PcG Complex, and Histone H3 K4/K27 Trimethylation', *Cell Stem Cell*, 2(2), pp. 160–169. doi: 10.1016/j.stem.2007.12.011.

Frauer, C. *et al.* (2011) 'Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain', *PLoS ONE*. Edited by S. Xu, 6(6), p. e21306. doi: 10.1371/journal.pone.0021306.

Freitag, M. and Selker, E. U. (2005) 'Controlling DNA methylation: many roads to one modification', *Current Opinion in Genetics & Development*, 15(2), pp. 191–199. doi: 10.1016/j.gde.2005.02.003.

Fuks, F. *et al.* (2000) 'DNA methyltransferase Dnmt1 associates with histone deacetylase activity', *Nature Genetics*, 24(1), pp. 88–91. doi: 10.1038/71750.

Fuks, F. *et al.* (2003) 'The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase.', *Nucleic acids research*. Oxford University Press, 31(9), pp. 2305–12. doi: 10.1093/nar/gkg332.

Galonska, C. *et al.* (2015) 'Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming', *Cell Stem Cell*, 17(4), pp. 462–470. doi: 10.1016/j.stem.2015.07.005.

Gao, Y. *et al.* (2013) 'Replacement of Oct4 by Tet1 during iPSC Induction Reveals an Important Role of DNA Methylation and Hydroxymethylation in Reprogramming', *Cell Stem Cell*, 12(4), pp. 453–469. doi: 10.1016/j.stem.2013.02.005.

Gao, Z. *et al.* (2014) 'An AUTS2–Polycomb complex activates gene expression in the CNS', *Nature*, 516(7531), pp. 349–354. doi: 10.1038/nature13921.

Gapp, K. *et al.* (2014) 'Epigenetic regulation in neurodevelopment and neurodegenerative diseases', *Neuroscience*, 264, pp. 99–111. doi: 10.1016/j.neuroscience.2012.11.040.

Garcia-Gonzalo, F. R. and Izpisúa Belmonte, J. C. (2008) 'Albumin-associated lipids regulate human embryonic stem cell self-renewal.', *PloS one*. Public Library of Science, 3(1), p. e1384. doi: 10.1371/journal.pone.0001384.

Garcia-Perez, J. L., Widmann, T. J. and Adams, I. R. (2016) 'The impact of transposable elements on mammalian development', *Development*, 143(22), pp. 4101–4114. doi: 10.1242/dev.132639.

Gaudet, F. *et al.* (2003) 'Induction of Tumors in Mice by Genomic Hypomethylation', *Science*, 300(5618), pp. 489–492. doi: 10.1126/science.1083558.

Gaudet, F. *et al.* (2004) 'Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing.', *Molecular and cellular biology*, 24(4), pp. 1640–8. doi: 10.1128/mcb.24.4.1640-1648.2004.

Gehring, M., Reik, W. and Henikoff, S. (2009) 'DNA demethylation by DNA repair', *Trends in Genetics*, 25(2), pp. 82–90. doi: 10.1016/j.tig.2008.12.001.

Geiman, T. M. and Robertson, K. D. (2002) 'Chromatin remodeling, histone modifications, and DNA methylation?how does it all fit together?', *Journal of Cellular Biochemistry*, 87(2), pp. 117–125. doi: 10.1002/jcb.10286.

Georgel, P. T. *et al.* (2003) 'Chromatin Compaction by Human MeCP2', *Journal of Biological Chemistry*, 278(34), pp. 32181–32188. doi: 10.1074/jbc.M305308200.

Glickman, J. F., Pavlovich, J. G. and Reich, N. O. (1997) 'Peptide mapping of the murine DNA methyltransferase reveals a major phosphorylation site and the start of translation.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 272(28), pp. 17851–7. doi: 10.1074/jbc.272.28.17851.

Goll, M. G. and Bestor, T. H. (2005) 'Eukaryotic Cytosine Methyltransferases', *Annual Review of Biochemistry*. Annual Reviews, 74(1), pp. 481–514. doi: 10.1146/annurev.biochem.74.010904.153721.

Gonzalo, S. *et al.* (2006) 'DNA methyltransferases control telomere length and telomere recombination in mammalian cells', *Nature Cell Biology*, 8(4), pp. 416–424. doi: 10.1038/ncb1386.

Goyal, R. *et al.* (2007) 'Phosphorylation of Serine-515 Activates the Mammalian Maintenance Methyltransferase Dnmt1', *Epigenetics*, 2(3), pp. 155–160. doi: 10.4161/epi.2.3.4768.

Goyal, R., Reinhardt, R. and Jeltsch, A. (2006) 'Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase', *Nucleic Acids Research*, 34(4), pp. 1182–1188. doi: 10.1093/nar/gkl002.

Grabole, N. *et al.* (2013) 'Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation', *EMBO reports*, 14(7), pp. 629–637. doi: 10.1038/embor.2013.67.

Greenberg, M. V. C. *et al.* (2017) 'Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth', *Nature Genetics*, 49(1), pp. 110–118. doi: 10.1038/ng.3718.

Gu, B. *et al.* (2017) 'AIRE is a critical spindle-associated protein in embryonic stem cells', *eLife*, 6. doi: 10.7554/eLife.28131.

Gu, T.-P. *et al.* (2011) 'The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes', *Nature*, 477(7366), pp. 606–610. doi: 10.1038/nature10443.

Guo, F. *et al.* (2014) 'Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote', *Cell Stem Cell*, 15(4), pp. 447–459. doi: 10.1016/j.stem.2014.08.003.

Guo, G. *et al.* (2016) 'Serum-based culture conditions provoke gene expression variability in mouse embryonic stem cells as revealed by single cell analysis', *Cell reports*. NIH Public Access, 14(4), p. 956. doi: 10.1016/J.CELREP.2015.12.089.

Guy, J. *et al.* (2001) 'A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome', *Nature Genetics*, 27(3), pp. 322–326. doi: 10.1038/85899.

Guy, J. *et al.* (2007) 'Reversal of Neurological Defects in a Mouse Model of Rett Syndrome', *Science*, 315(5815), pp. 1143–1147. doi: 10.1126/science.1138389.

Habibi, E. *et al.* (2013a) 'Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells', *Cell Stem Cell*, 13(3), pp. 360–369. doi: 10.1016/j.stem.2013.06.002.

Habibi, E. *et al.* (2013b) 'Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells', *Cell Stem Cell*, 13(3), pp. 360–369. doi: 10.1016/j.stem.2013.06.002.

Habibi, E. *et al.* (2013c) 'Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells', *Cell Stem Cell*, 13(3), pp. 360–369. doi: 10.1016/j.stem.2013.06.002.

Hackett, Jamie A *et al.* (2012) 'Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline.', *Development (Cambridge, England)*, 139(19), pp. 3623–32. doi: 10.1242/dev.081661.

Hackett, J. A. *et al.* (2012a) 'Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline', *Development*, 139(19), pp. 3623–3632. doi: 10.1242/dev.081661.

Hackett, J. A. *et al.* (2012b) 'Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline', *Development*, 139(19), pp. 3623–3632. doi: 10.1242/dev.081661.

Hackett, J. A. *et al.* (2013) 'Synergistic Mechanisms of DNA Demethylation during Transition to Ground-State Pluripotency', *Stem Cell Reports*, 1(6), pp. 518–531. doi: 10.1016/j.stemcr.2013.11.010.

Hackett, Jamie A *et al.* (2017) 'Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum.', *Stem cell reports*. Elsevier, 8(6), pp. 1645–1658. doi: 10.1016/j.stemcr.2017.05.014.

Hackett, Jamie A. *et al.* (2017) 'Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum', *Stem Cell Reports*, 8(6), pp. 1645–1658. doi: 10.1016/j.stemcr.2017.05.014.

Hajkova, P. *et al.* (2010) 'Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway', *Science*, 329(5987), pp. 78–82. doi: 10.1126/science.1187945.

Han, X. *et al.* (2008) 'Mass spectrometry for proteomics.', *Current opinion in chemical biology*. NIH Public Access, 12(5), pp. 483–90. doi: 10.1016/j.cbpa.2008.07.024.

Han, Y. *et al.* (2017) 'Lsh/HELLS regulates self-renewal/proliferation of neural stem/progenitor cells', *Scientific Reports*, 7(1), p. 1136. doi: 10.1038/s41598-017-00804-6.

Hanna, J. *et al.* (2009) 'Metastable Pluripotent States in NOD-Mouse-Derived ESCs', *Cell Stem Cell*, 4(6), pp. 513–524. doi: 10.1016/j.stem.2009.04.015.

Hartl, D. *et al.* (2019) 'CG dinucleotides enhance promoter activity independent of DNA methylation.', *Genome research*. Cold Spring Harbor Laboratory Press, 29(4), pp. 554–563. doi: 10.1101/gr.241653.118.

Hashimoto, H. *et al.* (2008) 'The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix', *Nature*. Nature Publishing Group, 455(7214), pp. 826–829. doi: 10.1038/nature07280.

Hashimoto, H. *et al.* (2012) 'Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation', *Nucleic Acids Research*, 40(11), pp. 4841–4849. doi: 10.1093/nar/gks155.

Hata, K. *et al.* (2002) 'Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice.', *Development (Cambridge, England)*, 129(8), pp. 1983–93.

Hayashi, K. *et al.* (2008) 'Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States', *Cell Stem Cell*, 3(4), pp. 391–401. doi: 10.1016/j.stem.2008.07.027.

He, Y.-F. *et al.* (2011) 'Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA', *Science*, 333(6047), pp. 1303–1307. doi: 10.1126/science.1210944.

Hendrich, B. and Tweedie, S. (2003) 'The methyl-CpG binding domain and the evolving role of DNA methylation in animals', *Trends in Genetics*, 19(5), pp. 269–277. doi: 10.1016/S0168-9525(03)00080-5.

Hermann, A., Goyal, R. and Jeltsch, A. (2004) 'The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites', *Journal of Biological Chemistry*, 279(46), pp. 48350–48359. doi: 10.1074/jbc.M403427200.

Hervouet, E. *et al.* (2018) 'Specific or not specific recruitment of DNMTs for DNA methylation, an epigenetic dilemma', *Clinical Epigenetics*, 10(1), p. 17. doi: 10.1186/s13148-018-0450-y.

Hirasawa, R. *et al.* (2008) 'Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development', *Genes & Development*, 22(12), pp. 1607–1616. doi: 10.1101/gad.1667008.

Holliday, R. and Pugh, J. E. (1975) 'DNA modification mechanisms and gene activity during development.', *Science (New York, N.Y.)*, 187(4173), pp. 226–32. doi: 10.1126/science.173.3994.293.

Horii, T. *et al.* (2003) 'Serum-free culture of murine primordial germ cells and embryonic germ cells.', *Theriogenology*, 59(5–6), pp. 1257–64. doi: 10.1016/S0093-691X(02)01166-4.

Horvath, S. (2013) 'DNA methylation age of human tissues and cell types', *Genome Biology*, 14(10), p. R115. doi: 10.1186/gb-2013-14-10-r115.

Huang, Y. *et al.* (2017) 'Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition', *eLife*, 6. doi: 10.7554/eLife.22345.

Hussein, S. M. I. *et al.* (2014) 'Genome-wide characterization of the routes to pluripotency', *Nature*, 516(7530), pp. 198–206. doi: 10.1038/nature14046.

ter Huurne, M. *et al.* (2017) 'Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency', *Cell Stem Cell*, 21(4), pp. 449-455.e4. doi: 10.1016/j.stem.2017.09.004.

Illingworth, R. *et al.* (2008) 'A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci', *PLoS Biology*. Edited by E. T. Liu, 6(1), p. e22. doi: 10.1371/journal.pbio.0060022.

Illingworth, R. S. *et al.* (2010) 'Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome', *PLoS Genetics*. Edited by W. Reik, 6(9), p. e1001134. doi: 10.1371/journal.pgen.1001134.

Illingworth, R. S. and Bird, A. P. (2009) 'CpG islands - "A rough guide"', *FEBS Letters*, 583(11), pp. 1713–1720. doi: 10.1016/j.febslet.2009.04.012.

Imamura, M. *et al.* (2006) 'Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells.', *BMC developmental biology*. BioMed Central, 6, p. 34. doi: 10.1186/1471-213X-6-34.

Inoue, A., Matoba, S. and Zhang, Y. (2012) 'Transcriptional activation of transposable elements in mouse zygotes is independent of Tet3-mediated 5-methylcytosine oxidation', *Cell Research*, 22(12), pp. 1640–1649. doi: 10.1038/cr.2012.160.

International Human Genome Sequencing Consortium (2001) 'Initial sequencing and analysis of the human genome', *Nature*. Nature Publishing Group, 409(6822), pp. 860–921. doi: 10.1038/35057062.

Isagawa, T. *et al.* (2011) 'DNA Methylation Profiling of Embryonic Stem Cell Differentiation into the Three Germ Layers', *PLoS ONE*. Edited by A. Imhof, 6(10), p. e26052. doi: 10.1371/journal.pone.0026052.

Ishida-Yamamoto, A. *et al.* (2000) 'Decreased Deiminated Keratin K1 in Psoriatic Hyperproliferative Epidermis', *Journal of Investigative Dermatology*, 114(4), pp. 701–705. doi: 10.1046/j.1523-1747.2000.00936.x.

Ishiyama, S. *et al.* (2017) 'Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance', *Molecular Cell*, 68(2), pp. 350-360.e7. doi: 10.1016/j.molcel.2017.09.037.

Jachowicz, J. W. *et al.* (2017) 'LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo', *Nature Genetics*, 49(10), pp. 1502–1510. doi: 10.1038/ng.3945.

Jackson-Grusby, L. *et al.* (2001) 'Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation', *Nature Genetics*, 27(1), pp. 31–39. doi: 10.1038/83730.

Jackson, M. *et al.* (2004a) 'Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells', *Molecular and Cellular Biology*, 24(20), pp. 8862–8871. doi: 10.1128/MCB.24.20.8862-8871.2004.

Jackson, M. *et al.* (2004b) 'Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells', *Molecular and Cellular Biology*, 24(20), pp. 8862–8871. doi: 10.1128/MCB.24.20.8862-8871.2004.

Jeltsch, A. (2006) 'Molecular enzymology of mammalian DNA methyltransferases', in *Current Topics in Microbiology and Immunology*, pp. 203–225. doi: 10.1007/3-540-31390-7_7.

Jeltsch, A. *et al.* (2017) 'Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation', *RNA Biology*, 14(9), pp. 1108–1123. doi: 10.1080/15476286.2016.1191737.

Jeltsch, A. and Jurkowska, R. Z. (2016) 'Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm.', *Nucleic acids research*. Oxford University Press, 44(18), pp. 8556–8575. doi: 10.1093/nar/gkw723.

Jenness, C. *et al.* (2018) 'HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome', *Proceedings of the National Academy of Sciences*, 115(5), pp. E876–E885. doi: 10.1073/pnas.1717509115.

Jia, J. *et al.* (2012) 'Regulation of Pluripotency and Self- Renewal of ESCs through Epigenetic-Threshold Modulation and mRNA Pruning', *Cell*, 151(3), pp. 576–589. doi: 10.1016/j.cell.2012.09.023.

Jiang, Y. L. *et al.* (2005) 'DNMT3B mutations and DNA methylation defect define two types of ICF syndrome', *Human Mutation*, 25(1), pp. 56–63. doi: 10.1002/humu.20113.

Jin, B. *et al.* (2008) 'DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function', *Human Molecular Genetics*, 17(5), pp. 690–709. doi: 10.1093/hmg/ddm341.

Jones, P. A. and Baylin, S. B. (2002) 'The fundamental role of epigenetic events in cancer', *Nature Reviews Genetics*, 3(6), pp. 415–428. doi: 10.1038/nrg816.

Jones, P. A., Issa, J.-P. J. and Baylin, S. (2016) 'Targeting the cancer epigenome for therapy', *Nature Reviews Genetics*, 17(10), pp. 630–641. doi: 10.1038/nrg.2016.93.

Jordan, I. K. *et al.* (2003) 'Origin of a substantial fraction of human regulatory sequences from transposable elements.', *Trends in genetics : TIG*, 19(2), pp. 68–72. doi:10.1016/S0168-9525(02)00006-9.

Jung, H.-J. *et al.* (2017) 'The Ubiquitin-like with PHD and Ring Finger Domains 1 (UHRF1)/DNA Methyltransferase 1 (DNMT1) Axis Is a Primary Regulator of Cell Senescence', *Journal of Biological Chemistry*, 292(9), pp. 3729–3739. doi: 10.1074/jbc.M116.750539.

Jurkowska, R. Z., Jurkowski, T. P. and Jeltsch, A. (2011) 'Structure and Function of Mammalian DNA Methyltransferases', *ChemBioChem*, 12(2), pp. 206–222. doi: 10.1002/cbic.201000195.

Jurkowski, T. P. and Jeltsch, A. (2011) 'On the Evolutionary Origin of Eukaryotic DNA Methyltransferases and Dnmt2', *PLoS ONE*. Edited by F. Lyko. Public Library of Science, 6(11), p. e28104. doi: 10.1371/journal.pone.0028104.

Kalscheuer, V. M. *et al.* (2007) 'Mutations in autism susceptibility candidate 2 (AUTS2) in patients with mental retardation', *Human Genetics*, 121(3–4), pp. 501–509. doi: 10.1007/s00439-006-0284-0.

Kan, R. *et al.* (2012) 'Potential role for PADI-mediated histone citrullination in preimplantation development', *BMC Developmental Biology*, 12(1), p. 19. doi: 10.1186/1471-213X-12-19.

Kanada, K. *et al.* (2017) 'Conserved threonine 1505 in the catalytic domain stabilizes mouse DNA methyltransferase 1', *The Journal of Biochemistry*, 162(4), pp. 271–278. doi: 10.1093/jb/mvx024.

Kaneda, M. *et al.* (2004) 'Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting', *Nature*, 429(6994), pp. 900–903. doi: 10.1038/nature02633.

Kar, S. *et al.* (2012) 'An insight into the various regulatory mechanisms modulating human DNA methyltransferase 1 stability and function', *Epigenetics*, 7(9), pp. 994–1007. doi: 10.4161/epi.21568.

Karimi, Mohammad M *et al.* (2011) 'DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs', *Stem Cell*, 8, pp. 676–687. doi: 10.1016/j.stem.2011.04.004.

Karimi, Mohammad M. *et al.* (2011) 'DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs', *Cell Stem Cell*, 8(6), pp. 676–687. doi: 10.1016/j.stem.2011.04.004.

Kigami, D. *et al.* (2003) 'MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos1', *Biology of Reproduction*. Narnia, 68(2), pp. 651–654. doi: 10.1095/biolreprod.102.007906.

Kim, J.-A. and Yeom, Y. Il (2018) 'Metabolic Signaling to Epigenetic Alterations in Cancer.', *Biomolecules & therapeutics*. Korean Society of Applied Pharmacology, 26(1), pp. 69–80. doi: 10.4062/biomolther.2017.185.

Kim, K.-Y. *et al.* (2018) 'Uhrf1 regulates active transcriptional marks at bivalent domains in pluripotent stem cells through Setd1a', *Nature Communications*. Nature Publishing Group, 9(1), p. 2583. doi: 10.1038/s41467-018-04818-0.

Kim, T.-W. *et al.* (2019) 'Application of TurboID-mediated proximity labeling for mapping a GSK3 kinase signaling network in Arabidopsis', *bioRxiv*. Cold Spring Harbor Laboratory, p. 636324. doi: 10.1101/636324.

King, Andrew D *et al.* (2016) 'Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells.', *Cell reports*. Elsevier, 17(1), pp. 289–302. doi: 10.1016/j.celrep.2016.08.083.

King, Andrew D. *et al.* (2016) 'Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells', *Cell Reports*, 17(1), pp. 289–302. doi: 10.1016/j.celrep.2016.08.083.

Kiyonari, H. *et al.* (2010) 'Three inhibitors of FGF receptor, ERK, and GSK3 establishes germline-competent embryonic stem cells of C57BL/6N mouse strain with high efficiency and stability', *genesis*, 48(5), p. NA-NA. doi: 10.1002/dvg.20614.

Knorre, D. G., Kudryashova, N. V and Godovikova, T. S. (2009) 'Chemical and functional aspects of posttranslational modification of proteins.', *Acta naturae*, 1(3), pp. 29–51.doi:10.32607/20758251-2009-1-3-29-51

Kohli, R. M. and Zhang, Y. (2013) 'TET enzymes, TDG and the dynamics of DNA demethylation', *Nature*, 502(7472), pp. 472–479. doi: 10.1038/nature12750.

Komander, D. and Rape, M. (2012) 'The Ubiquitin Code', *Annual Review of Biochemistry*, 81(1), pp. 203–229. doi: 10.1146/annurev-biochem-060310-170328.

Kooistra, S. M. *et al.* (2010) 'Undifferentiated Embryonic Cell Transcription Factor 1 Regulates ESC Chromatin Organization and Gene Expression', *STEM CELLS*, 28(10), pp. 1703–1714. doi: 10.1002/stem.497.

Kriaucionis, S. and Heintz, N. (2009) 'The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain', *Science*, 324(5929), pp. 929–930. doi: 10.1126/science.1169786.

Kribelbauer, J. F. *et al.* (2017) 'Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes', *Cell Reports*, 19(11), pp. 2383–2395. doi: 10.1016/j.celrep.2017.05.069.

Kumar, S. *et al.* (1994) 'The DNA (cytosine-5) methyltransferases', *Nucleic Acids Research*, 22(1), pp. 1–10. doi: 10.1093/nar/22.1.1.

Kunath, T. *et al.* (2007) 'FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment', *Development*, 134(16), pp. 2895–2902. doi: 10.1242/dev.02880.

Kunert, N. *et al.* (2003) 'A Dnmt2-like protein mediates DNA methylation in Drosophila', *Development*, 130(21), pp. 5083–5090. doi: 10.1242/dev.00716.

Kurihara, Y. *et al.* (2008) 'Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1', *Developmental Biology*, 313(1), pp. 335–346. doi: 10.1016/j.ydbio.2007.10.033.

Ladstätter, S. and Tachibana-Konwalski, K. (2016) 'A Surveillance Mechanism Ensures Repair of DNA Lesions during Zygotic Reprogramming.', *Cell*. Elsevier, 167(7), pp. 1774-1787.e13. doi: 10.1016/j.cell.2016.11.009.

Lambert, J.-P. *et al.* (2015) 'Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes', *Journal of Proteomics*, 118, pp. 81–94. doi: 10.1016/j.jprot.2014.09.011.

Lavoie, G. and St-Pierre, Y. (2011) 'Phosphorylation of human DNMT1: Implication of cyclin-dependent kinases', *Biochemical and Biophysical Research Communications*, 409(2), pp. 187–192. doi: 10.1016/j.bbrc.2011.04.115.

Lee, B. and Muller, M. T. (2009) 'SUMOylation enhances DNA methyltransferase 1 activity', *Biochemical Journal*, 421(3), pp. 449–461. doi: 10.1042/BJ20090142.

Lee, D.-S. *et al.* (2014) 'An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator', *Nature Communications*, 5(1), p. 5619. doi: 10.1038/ncomms6619.

Lei, H. *et al.* (1996) 'De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells.', *Development (Cambridge, England)*, 122(10), pp. 3195–205.

Leighton, P. A. *et al.* (1995) 'Disruption of imprinting caused by deletion of the H19 gene region in mice', *Nature*, 375(6526), pp. 34–39. doi: 10.1038/375034a0.

Leitch, H. G. *et al.* (2010) 'Embryonic germ cells from mice and rats exhibit properties consistent with a generic pluripotent ground state', *Development*, 137(14), pp. 2279–2287. doi: 10.1242/dev.050427.

Leitch, H. G. *et al.* (2013) 'Naive pluripotency is associated with global DNA hypomethylation', *Nature Structural & Molecular Biology*, 20(3), pp. 311–316. doi: 10.1038/nsmb.2510.

Leonard, H., Cobb, S. and Downs, J. (2017) 'Clinical and biological progress over 50 years in Rett syndrome', *Nature Reviews Neurology*. Nature Publishing Group, 13(1), pp. 37–51. doi: 10.1038/nrneurol.2016.186.

Leung, D. C. *et al.* (2011) 'Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing', *Proceedings of the National Academy of Sciences*, 108(14), pp. 5718–5723. doi: 10.1073/pnas.1014660108.

Li, C. *et al.* (2018) 'DNA methylation reprogramming of functional elements during mammalian embryonic development', *Cell Discovery*. Nature Publishing Group, 4(1), p. 41. doi: 10.1038/s41421-018-0039-9.

Li, E., Bestor, T. H. and Jaenisch, R. (1992) 'Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.', *Cell*, 69(6), pp. 915–26. doi: 10.1016/0092-8674(92)90611-f.

Li, H. *et al.* (2006) 'The Histone Methyltransferase SETDB1 and the DNA Methyltransferase DNMT3A Interact Directly and Localize to Promoters Silenced in Cancer Cells', *Journal of Biological Chemistry*, 281(28), pp. 19489–19500. doi: 10.1074/jbc.M513249200.

Li, P. *et al.* (2010) 'PAD4 is essential for antibacterial innate immunity mediated by neutrophil extracellular traps', *The Journal of Experimental Medicine*, 207(9), pp. 1853–1862. doi: 10.1084/jem.20100239.

Li, T. *et al.* (2018) 'Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation', *Nucleic Acids Research*, 46(6), pp. 3218–3231. doi: 10.1093/nar/gky104.

Li, X. *et al.* (2008) 'A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints', *Developmental Cell*, 15(4), pp. 547–557. doi: 10.1016/j.devcel.2008.08.014.

Li, X. *et al.* (2018) 'BRD4 Promotes DNA Repair and Mediates the Formation of TMPRSS2-ERG Gene Rearrangements in Prostate Cancer.', *Cell reports*. NIH Public Access, 22(3), pp. 796–808. doi: 10.1016/j.celrep.2017.12.078.

Li, Y. *et al.* (2018) 'Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1', *Nature*, 564(7734), pp. 136–140. doi: 10.1038/s41586-018-0751-5.

Li, Y. and Tollefsbol, T. O. (2011) 'DNA Methylation Detection: Bisulfite Genomic Sequencing Analysis', in *Methods in molecular biology (Clifton, N.J.)*, pp. 11–21. doi: 10.1007/978-1-61779-316-5_2.

Li, Z., Dai, H., Martos, Suzanne N, *et al.* (2015) 'Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells.', *Genome biology*, 16, p. 115. doi: 10.1186/s13059-015-0685-2.

Li, Z., Dai, H., Martos, Suzanne N., *et al.* (2015) 'Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells', *Genome Biology*, 16(1), p. 115. doi: 10.1186/s13059-015-0685-2.

Liang, G. *et al.* (2002) 'Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements.', *Molecular and cellular biology*, 22(2), pp. 480–91. doi: 10.1128/mcb.22.2.480-491.2002.

Liao, J. *et al.* (2015) 'Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells', *Nature Genetics*, 47(5), pp. 469–478. doi: 10.1038/ng.3258.

Lienert, F. *et al.* (2011) 'Identification of genetic elements that autonomously determine DNA methylation states', *Nature Genetics*, 43(11), pp. 0–9. doi: 10.1038/ng.946.

Liu, G. *et al.* (2018) 'Inherited DNA methylation primes the establishment of accessible chromatin during genome activation.', *Genome research*. Cold Spring Harbor Laboratory Press, 28(7), pp. 998–1007. doi: 10.1101/gr.228833.117.

Lorincz, M. C. *et al.* (2002) 'DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b-independent de novo methylation.', *Molecular and cellular biology*, 22(21), pp. 7572–80. doi: 10.1128/mcb.22.21.7572-7580.2002.

Lu, L.-Y. *et al.* (2015a) 'Topoisomerase II Regulates the Maintenance of DNA Methylation', *Journal of Biological Chemistry*, 290(2), pp. 851–860. doi: 10.1074/jbc.M114.611509.

Lu, L.-Y. *et al.* (2015b) 'Topoisomerase II Regulates the Maintenance of DNA Methylation', *Journal of Biological Chemistry*, 290(2), pp. 851–860. doi: 10.1074/jbc.M114.611509.

Lyko, F., Ramsahoye, B. H. and Jaenisch, R. (2000) 'DNA methylation in Drosophila melanogaster', *Nature*. Nature Publishing Group, 408(6812), pp. 538–540. doi: 10.1038/35046205.

Lyons, D. B. and Zilberman, D. (2017) 'DDM1 and Lsh remodelers allow methylation of DNA wrapped in nucleosomes', *eLife*, 6. doi: 10.7554/eLife.30674.

Ma, H. *et al.* (2012) 'M phase phosphorylation of the epigenetic regulator UHRF1 regulates its physical association with the deubiquitylase USP7 and stability', *Proceedings of the National Academy of Sciences*, 109(13), pp. 4828–4833. doi: 10.1073/pnas.1116349109.

Macfarlan, T. S. *et al.* (2012) 'Embryonic stem cell potency fluctuates with endogenous retrovirus activity', *Nature*, 487(7405), pp. 57–63. doi: 10.1038/nature11244.

Mackay, A. *et al.* (2009) 'A high-resolution integrated analysis of genetic and expression profiles of breast cancer cell lines', *Breast Cancer Research and Treatment*, 118(3), pp. 481–498. doi: 10.1007/s10549-008-0296-7.

Mackay, D. J. G. *et al.* (2008) 'Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57', *Nature Genetics*, 40(8), pp. 949–951. doi: 10.1038/ng.187.

Maenohara, S. *et al.* (2017) 'Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos', *PLOS Genetics*. Edited by P. E. Cohen. Public Library of Science, 13(10), p. e1007042. doi: 10.1371/journal.pgen.1007042.

Mandemaker, I. K. *et al.* (2017) 'DNA damage-induced histone H1 ubiquitylation is mediated by HUWE1 and stimulates the RNF8-RNF168 pathway', *Scientific Reports*, 7(1), p. 15353. doi: 10.1038/s41598-017-15194-y.

Marchal, C. and Miotto, B. (2015) 'Emerging Concept in DNA Methylation: Role of Transcription Factors in Shaping DNA Methylation Patterns', *Journal of Cellular Physiology*, 230(4), pp. 743–751. doi: 10.1002/jcp.24836.

Marin, T. L. *et al.* (2017) 'AMPK promotes mitochondrial biogenesis and function by phosphorylating the epigenetic factors DNMT1, RBBP7, and HAT1', *Science Signaling*, 10(464), p. eaaf7478. doi: 10.1126/scisignal.aaf7478.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., *et al.* (2012) 'The Transcriptional and Epigenomic Foundations of Ground State Pluripotency', *Cell*, 149(3), pp. 590–604. doi: 10.1016/j.cell.2012.03.026.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, a., *et al.* (2012) 'The Transcriptional and Epigenomic Foundations of Ground State Pluripotency', *Cell*, 149(3), pp. 590–604. doi: 10.1016/j.cell.2012.03.026.

Marks, H. and Stunnenberg, H. G. (2014) 'Transcription regulation and chromatin structure in the pluripotent ground state', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1839(3), pp. 129–137. doi: 10.1016/j.bbagrm.2013.09.005.

Martello, G. and Smith, A. (2014) 'The Nature of Embryonic Stem Cells', *Annual Review of Cell and Developmental Biology*. Annual Reviews , 30(1), pp. 647–675. doi: 10.1146/annurev-cellbio-100913-013116.

Martin-Trujillo, A. *et al.* (2017) 'Copy number rather than epigenetic alterations are the major dictator of imprinted methylation in tumors', *Nature Communications*, 8(1), p. 467. doi: 10.1038/s41467-017-00639-9.

Matsui, T. *et al.* (2010) 'Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET', *Nature*, 464(7290), pp. 927–931. doi: 10.1038/nature08858.

Maurano, M. T. *et al.* (2015) 'Role of DNA Methylation in Modulating Transcription Factor Occupancy', *Cell Reports*, 12(7), pp. 1184–1195. doi: 10.1016/j.celrep.2015.07.024.

McElwee, J. L. *et al.* (2012) 'Identification of PADI2 as a potential breast cancer biomarker and therapeutic target', *BMC Cancer*, 12(1), p. 500. doi: 10.1186/1471-2407-12-500.

McGraw, C. M., Samaco, R. C. and Zoghbi, H. Y. (2011) 'Adult Neural Function Requires MeCP2', *Science*, 333(6039), pp. 186–186. doi: 10.1126/science.1206593.

McGraw, S. *et al.* (2013) 'Loss of DNMT1o Disrupts Imprinted X Chromosome Inactivation and Accentuates Placental Defects in Females', *PLoS Genetics*. Edited by J. T. Lee, 9(11), p. e1003873. doi: 10.1371/journal.pgen.1003873.

Meissner, A., Mikkelsen, Tarjei S, *et al.* (2008) 'Genome-scale DNA methylation maps of pluripotent and differentiated cells.', *Nature*, 454(7205), pp. 766–70. doi: 10.1038/nature07107.

Meissner, A., Mikkelsen, Tarjei S., *et al.* (2008) 'Genome-scale DNA methylation maps of pluripotent and differentiated cells', *Nature*, 454(7205), pp. 766–770. doi: 10.1038/nature07107.

Meissner, A. (2010) 'Epigenetic modifications in pluripotent and differentiated cells', *Nature Biotechnology*. Nature Publishing Group, 28(10), pp. 1079–1088. doi: 10.1038/nbt.1684.

Meredith, G. D. *et al.* (2015) 'Glycogen synthase kinase-3 (Gsk-3) plays a fundamental role in maintaining DNA methylation at imprinted loci in mouse embryonic stem cells', *Molecular Biology of the Cell*. Edited by J. Chernoff, 26(11), pp. 2139–2150. doi: 10.1091/mbc.E15-01-0013.

Mikkelsen, T. S. *et al.* (2007) 'Genome-wide maps of chromatin state in pluripotent and lineage-committed cells', *Nature*, 448(7153), pp. 553–560. doi: 10.1038/nature06008.

Mikkelsen, T. S. *et al.* (2008) 'Dissecting direct reprogramming through integrative genomic analysis.', *Nature*. Nature Publishing Group, 454(7200), pp. 49–55. doi: 10.1038/nature07056.

Monderer-Rothkoff, G. *et al.* (2019) 'AUTS2 isoforms control neuronal differentiation', *Molecular Psychiatry*. doi: 10.1038/s41380-019-0409-1.

Moore, L. D., Le, T. and Fan, G. (2012) 'DNA Methylation and Its Basic Function', *Neuropsychopharmacology*. American College of Neuropsychopharmacology, 38(1), pp. 23–38. doi: 10.1038/npp.2012.112.

Morgan, H. D. *et al.* (2004) 'Activation-induced Cytidine Deaminase Deaminates 5-Methylcytosine in DNA and Is Expressed in Pluripotent Tissues', *Journal of Biological Chemistry*, 279(50), pp. 52353–52360. doi: 10.1074/jbc.M407695200.

Morgani, S. M. *et al.* (2013) 'Totipotent Embryonic Stem Cells Arise in Ground-State Culture Conditions', *Cell Reports*, 3(6), pp. 1945–1957. doi: 10.1016/j.celrep.2013.04.034.

Mouse Genome Sequencing Consortium *et al.* (2002) 'Initial sequencing and comparative analysis of the mouse genome', *Nature*, 420(6915), pp. 520–562. doi: 10.1038/nature01262.

Mulholland, C. B. *et al.* (2015) 'A modular open platform for systematic functional studies under physiological conditions.', *Nucleic acids research*, 43(17), p. e112. doi: 10.1093/nar/gkv550.

Mulholland, C. B. *et al.* (2018) 'TET1 drives global DNA demethylation via DPPA3-mediated inhibition of maintenance methylation', *bioRxiv*. Cold Spring Harbor Laboratory, p. 321604. doi: 10.1101/321604.

Murphy, P. J. *et al.* (2018) 'Placeholder Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming', *Cell*, 172(5), pp. 993-1006.e13. doi: 10.1016/j.cell.2018.01.022.

Nabel, C. S. *et al.* (2012a) 'AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation.', *Nature chemical biology*. NIH Public Access, 8(9), pp. 751–8. doi: 10.1038/nchembio.1042.

Nabel, C. S. *et al.* (2012b) 'AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation.', *Nature chemical biology*. NIH Public Access, 8(9), pp. 751–8. doi: 10.1038/nchembio.1042.

Nakamura, T. *et al.* (2007) 'PGC7/Stella protects against DNA demethylation in early embryogenesis', *NATURE CELL BIOLOGY*, 9(1). doi: 10.1038/ncb1519.

Nakamura, T. *et al.* (2012a) 'PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos', *Nature*, 486(7403), pp. 415–419. doi: 10.1038/nature11093.

Nakamura, T. *et al.* (2012b) 'PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos', *Nature*, 486(7403), pp. 415–419. doi: 10.1038/nature11093.

Nakanishi, M. O. *et al.* (2012) 'Trophoblast-specific DNA methylation occurs after the segregation of the trophectoderm and inner cell mass in the mouse periimplantation embryo', *Epigenetics*, 7(2), pp. 173–182. doi: 10.4161/epi.7.2.18962.

Nakashima, K. *et al.* (2013) 'PAD4 regulates proliferation of multipotent haematopoietic cells by controlling c-myc expression', *Nature Communications*. Nature Publishing Group, 4(1), p. 1836. doi: 10.1038/ncomms2862.

Neri, F. *et al.* (2017) 'Intragenic DNA methylation prevents spurious transcription initiation', *Nature*, 543(7643), pp. 72–77. doi: 10.1038/nature21373.

Nestor, C. E. *et al.* (2015) 'Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems', *Genome Biology*, 16(1), p. 11. doi: 10.1186/s13059-014-0576-y.

Nichols, J. *et al.* (2009) 'Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo', *Development*, 136(19), pp. 3215–3222. doi: 10.1242/dev.038893.

Nichols, J. and Smith, A. (2009) 'Naive and Primed Pluripotent States', *Cell Stem Cell*, 4(6), pp. 487–492. doi: 10.1016/j.stem.2009.05.015.

Nishiyama, A. *et al.* (2013) 'Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication', *Nature*, 502(7470), pp. 249–253. doi: 10.1038/nature12488.

O'Neill, K. M. *et al.* (2018) 'Depletion of DNMT1 in differentiated human cells highlights key classes of sensitive genes and an interplay with polycomb repression', *Epigenetics & Chromatin*. BioMed Central, 11(1), p. 12. doi: 10.1186/s13072-018-0182-4.

Okano, M *et al.* (1999) 'DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.', *Cell*, 99(3), pp. 247–57. doi: 10.1016/s0092-8674(00)81656-6.

Okano, Masaki *et al.* (1999) 'DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development', *Cell*, 99, pp. 247–257.

Okano, M., Xie, S. and Li, E. (1998) 'Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases', *Nature Genetics*, 19(3), pp. 219–220. doi: 10.1038/890.

Okeoma, C. M. *et al.* (2007) 'APOBEC3 inhibits mouse mammary tumour virus replication in vivo', *Nature*, 445(7130), pp. 927–930. doi: 10.1038/nature05540.

Okuda, A. *et al.* (1998) 'UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells', *The EMBO Journal*, 17(7), pp. 2019–2032. doi: 10.1093/emboj/17.7.2019.

Ooi, S. K. T. *et al.* (2007) 'DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA', *Nature*, 448(7154), pp. 714–717. doi: 10.1038/nature05987.

Otani, J. *et al.* (2009) 'Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain', *EMBO reports*, 10(11), pp. 1235–1241. doi: 10.1038/embor.2009.218.

Pannell, D. *et al.* (2000) 'Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code', *The EMBO Journal*, 19(21), pp. 5884–5894. doi: 10.1093/emboj/19.21.5884.

Panning, B. and Jaenisch, R. (1996) 'DNA hypomethylation can activate Xist expression and silence X-linked genes.', *Genes & Development*, 10(16), pp. 1991–2002. doi: 10.1101/gad.10.16.1991.

Papait, R. *et al.* (2008) 'The PHD domain of Np95 (mUHRF1) is involved in large-scale reorganization of pericentromeric heterochromatin.', *Molecular biology of the cell*. American Society for Cell Biology, 19(8), pp. 3554–63. doi: 10.1091/mbc.e07-10-1059.

Peat, J. R. *et al.* (2014) 'Genome-wide Bisulfite Sequencing in Zygotes Identifies Demethylation Targets and Maps the Contribution of TET3 Oxidation', *Cell Reports*, 9(6), pp. 1990–2000. doi: 10.1016/j.celrep.2014.11.034.

Pedrali-Noy, G. and Weissbach, A. (1986) 'Mammalian DNA methyltransferases prefer poly(dI-dC) as substrate.', *The Journal of biological chemistry*, 261(17), pp. 7600–2.

Peng, L. *et al.* (2011) 'SIRT1 Deacetylates the DNA Methyltransferase 1 (DNMT1) Protein and Alters Its Activities', *Molecular and Cellular Biology*, 31(23), pp. 4720–4734. doi: 10.1128/MCB.06147-11.

Petell, C. J. *et al.* (2016) 'An epigenetic switch regulates *de novo* DNA methylation at a subset of pluripotency gene enhancers during embryonic stem cell differentiation', *Nucleic Acids Research*, 44(16), pp. 7605–7617. doi: 10.1093/nar/gkw426.

Pfaffeneder, T. *et al.* (2014) 'Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA', *Nature Chemical Biology*, 10(7), pp. 574–581. doi: 10.1038/nchembio.1532.

Phizicky, E. M. and Fields, S. (1995) 'Protein-protein interactions: methods for detection and analysis.', *Microbiological reviews*, 59(1), pp. 94–123.

Poh, W. J., Wee, C. P. P. and Gao, Z. (2016) 'DNA Methyltransferase Activity Assays: Advances and Challenges.', *Theranostics*. Ivyspring International Publisher, 6(3), pp. 369–91. doi: 10.7150/thno.13438.

Pradhan, M. *et al.* (2008) 'CXXC Domain of Human DNMT1 Is Essential for Enzymatic Activity', *Biochemistry*, 47(38), pp. 10000–10009. doi: 10.1021/bi8011725.

Qin, W. *et al.* (2015) 'DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination', *Cell Research*, 25(8), pp. 911–929. doi: 10.1038/cr.2015.72.

Qin, W., Leonhardt, H. and Spada, F. (2011) 'Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1', *Journal of Cellular Biochemistry*, 112(2), pp. 439–444. doi: 10.1002/jcb.22998.

Quenneville, S. *et al.* (2011) 'In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions', *Molecular Cell*, 44(3), pp. 361–372. doi: 10.1016/j.molcel.2011.08.032.

Rajakumara, E. *et al.* (2011) 'PHD Finger Recognition of Unmodified Histone H3R2 Links UHRF1 to Regulation of Euchromatic Gene Expression', *Molecular Cell*, 43(2), pp. 275–284. doi: 10.1016/j.molcel.2011.07.006.

Ratnam, S. *et al.* (2002) 'Dynamics of Dnmt1 Methyltransferase Expression and Intracellular Localization during Oogenesis and Preimplantation Development', *Developmental Biology*, 245(2), pp. 304–314. doi: 10.1006/dbio.2002.0628.

Ravichandran, M., Jurkowska, R. Z. and Jurkowski, T. P. (2018) 'Target specificity of mammalian DNA methylation and demethylation machinery.', *Organic & biomolecular chemistry*, 16(9), pp. 1419–1435. doi: 10.1039/c7ob02574b.

Reale, A. *et al.* (2005) 'Modulation of DNMT1 activity by ADP-ribose polymers', *Oncogene*, 24(1), pp. 13–19. doi: 10.1038/sj.onc.1208005.

Reddington, J. P. *et al.* (2013) 'Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes', *Genome Biology*. BioMed Central, 14(3), p. R25. doi: 10.1186/gb-2013-14-3-r25.

Reddington, J. P., Pennings, S. and Meehan, R. R. (2013) 'Non-canonical functions of the DNA methylome in gene regulation', *Biochemical Journal*, 451(1), pp. 13–23. doi: 10.1042/BJ20121585.

Riggs, A. D. (1975) 'X inactivation, differentiation, and DNA methylation', *Cytogenetic and Genome Research*, 14(1), pp. 9–25. doi: 10.1159/000130315.

Riso, V. *et al.* (2016) 'ZFP57 maintains the parent-of-origin-specific expression of the imprinted genes and differentially affects non-imprinted targets in mouse embryonic stem cells', *Nucleic Acids Research*, 44(17), pp. 8165–8178. doi: 10.1093/nar/gkw505.

Robertson, K. D. *et al.* (2000) 'DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters', *Nature Genetics*, 25(3), pp. 338–342. doi: 10.1038/77124.

Romar, G. A., Kupper, T. S. and Divito, S. J. (2016) 'Research Techniques Made Simple: Techniques to Assess Cell Proliferation', *Journal of Investigative Dermatology*, 136(1), pp. e1–e7. doi: 10.1016/j.jid.2015.11.020.

Rothbart, S. B. *et al.* (2012) 'Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation', *Nature Structural & Molecular Biology*, 19(11). doi: 10.1038/nsmb.2391.

Roulois, D. *et al.* (2015) 'DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts.', *Cell*. Elsevier, 162(5), pp. 961–73. doi: 10.1016/j.cell.2015.07.056.

Rountree, M. R., Bachman, K. E. and Baylin, S. B. (2000a) 'DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci', *Nature Genetics*, 25(3), pp. 269–277. doi: 10.1038/77023.

Rountree, M. R., Bachman, K. E. and Baylin, S. B. (2000b) 'DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci', *Nature Genetics*, 25(3), pp. 269–277. doi: 10.1038/77023.

Rountree, M. R., Bachman, K. E. and Baylin, S. B. (no date) 'DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci'.

Roux, K. J. *et al.* (2012) 'A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells', *The Journal of Cell Biology*, 196(6), pp. 801–810. doi: 10.1083/jcb.201112098.

Roux, K. J. *et al.* (2018) 'BioID: A Screen for Protein-Protein Interactions', *Current protocols in protein science*. NIH Public Access, 91, p. 19.23.1. doi: 10.1002/CPPS.51.

Sado, T. (2004) 'De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation', *Development*, 131(5), pp. 975–982. doi: 10.1242/dev.00995.

Sakaue, M. *et al.* (2010) 'DNA Methylation Is Dispensable for the Growth and Survival of the Extraembryonic Lineages', *Current Biology*, 20(16), pp. 1452–1457. doi: 10.1016/j.cub.2010.06.050.

Santos, F. *et al.* (2013) 'Active demethylation in mouse zygotes involves cytosine deamination and base excision repair', *Epigenetics & Chromatin*, 6(1), p. 39. doi: 10.1186/1756-8935-6-39.

Sasai, N. and Defossez, P.-A. (2009) 'Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes', *The International Journal of Developmental Biology*, 53(2–3), pp. 323–334. doi: 10.1387/ijdb.082652ns.

Schermelleh, L. *et al.* (2007) 'Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation.', *Nucleic acids research*. Oxford University Press, 35(13), pp. 4301–12. doi: 10.1093/nar/gkm432.

Schiesser, S. *et al.* (2013) 'Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine', *Journal of the American Chemical Society*, 135(39), pp. 14593–14599. doi: 10.1021/ja403229y.

Schlesinger, S. and Goff, S. P. (2015) 'Retroviral Transcriptional Regulation and Embryonic Stem Cells: War and Peace', *Molecular and Cellular Biology*, 35(5), pp. 770–777. doi: 10.1128/MCB.01293-14.

Schneider, K. *et al.* (2013) 'Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling', *Nucleic Acids Research*, 41(9), pp. 4860–4876. doi: 10.1093/nar/gkt191.

Schneider, Katrin *et al.* (2013) 'Dissection of cell cycle–dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling', *Nucleic Acids Research*, 41(9), pp. 4860–4876. doi: 10.1093/nar/gkt191.

Schoorlemmer, J. *et al.* (2014) 'Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency.', *Frontiers in oncology*. Frontiers Media SA, 4, p. 14. doi: 10.3389/fonc.2014.00014.

Schübeler, D. *et al.* (2000) 'Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation.', *Molecular and cellular biology*, 20(24), pp. 9103–12. doi: 10.1128/mcb.20.24.9103-9112.2000.

Sharif, J. *et al.* (2007) 'The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA', *Nature*, 450(7171), pp. 908–912. doi: 10.1038/nature06397.

Sharif, J. *et al.* (2016) 'Activation of Endogenous Retroviruses in Dnmt1 −/− ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA', *Cell Stem Cell*, 19(1), pp. 81–94. doi: 10.1016/j.stem.2016.03.013.

Sharma, P. *et al.* (2012) 'Citrullination of Histone H3 Interferes with HP1-Mediated Transcriptional Repression', *PLoS Genetics*. Edited by B. A. Sullivan, 8(9), p. e1002934. doi: 10.1371/journal.pgen.1002934.

Sharma, S. *et al.* (2012) 'Lysine methyltransferase G9a is not required for DNMT3A/3B anchoring to methylated nucleosomes and maintenance of DNA methylation in somatic cells', *Epigenetics & Chromatin*, 5(1), p. 3. doi: 10.1186/1756-8935-5-3.

Shen, L. *et al.* (2014) 'Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes', *Cell Stem Cell*, 15(4), pp. 459–471. doi: 10.1016/j.stem.2014.09.002.

Shukla, S. *et al.* (2011) 'CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing', *Nature*, 479(7371), pp. 74–79. doi: 10.1038/nature10442.

Silva, J. *et al.* (2008) 'Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition', *PLoS Biology*. Edited by M. A. Goodell. Public Library of Science, 6(10), p. e253. doi: 10.1371/journal.pbio.0060253.

Silva, J. and Smith, A. (2008) 'Capturing pluripotency.', *Cell*. Elsevier, 132(4), pp. 532–6. doi: 10.1016/j.cell.2008.02.006.

Sim, Y.-J. *et al.* (2017) '2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms', *Stem Cell Reports*, 8(5), pp. 1312–1328. doi: 10.1016/j.stemcr.2017.04.001.

Singer, Z. S. *et al.* (2014) 'Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells', *Molecular Cell*, 55(2), pp. 319–331. doi: 10.1016/j.molcel.2014.06.029.

Smallwood, A. *et al.* (2007) 'Functional cooperation between HP1 and DNMT1 mediates gene silencing', *Genes & Development*, 21(10), pp. 1169–1178. doi: 10.1101/gad.1536807.

Smallwood, S. A. *et al.* (2011) 'Dynamic CpG island methylation landscape in oocytes and preimplantation embryos', *Nature Genetics*, 43(8), pp. 811–814. doi: 10.1038/ng.864.

Smallwood, S. A. *et al.* (2014) 'Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity', *Nature Methods*. Nature Publishing Group, 11(8), pp. 817–820. doi: 10.1038/nmeth.3035.

Smallwood, S. A. and Kelsey, G. (2012) 'De novo DNA methylation: a germ cell perspective', *Trends in Genetics*. Elsevier, 28(1), pp. 33–42. doi: 10.1016/j.tig.2011.09.004.

Smith, A. (2017) 'Formative pluripotency: the executive phase in a developmental continuum.', *Development (Cambridge, England)*. Oxford University Press for The Company of Biologists Limited, 144(3), pp. 365–373. doi: 10.1242/dev.142679.

Smith, Z. D. *et al.* (2012a) 'A unique regulatory phase of DNA methylation in the early mammalian embryo', *Nature*, 484(7394), pp. 339–344. doi: 10.1038/nature10960.

Smith, Z. D. *et al.* (2012b) 'A unique regulatory phase of DNA methylation in the early mammalian embryo', *Nature*, 484(7394), pp. 339–344. doi: 10.1038/nature10960.

Smith, Z. D. *et al.* (2014) 'DNA methylation dynamics of the human preimplantation embryo', *Nature*. Nature Publishing Group, 511(7511), pp. 611–615. doi: 10.1038/nature13581.

Smith, Z. D. and Meissner, A. (2013) 'DNA methylation: roles in mammalian development.', *Nature reviews. Genetics*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 14(3), pp. 204–20. doi: 10.1038/nrg3354.

Song, J. *et al.* (2011) 'Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation', *Science*, 331(6020), pp. 1036–1040. doi: 10.1126/science.1195380.

Song, J. *et al.* (2012) 'Structure-Based Mechanistic Insights into DNMT1-Mediated Maintenance DNA Methylation', *Science*, 335(6069), pp. 709–712. doi: 10.1126/science.1214453.

Sørensen, A. L. *et al.* (2010) 'Lineage-Specific Promoter DNA Methylation Patterns Segregate Adult Progenitor Cell Types', *Stem Cells and Development*, 19(8), pp. 1257–1266. doi: 10.1089/scd.2009.0309.

Spada, F. *et al.* (2007) 'DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells', *Journal of Cell Biology*. doi: 10.1083/jcb.200610062.

Spitz, F. and Furlong, E. E. M. (2012) 'Transcription factors: from enhancer binding to developmental control', *Nature Reviews Genetics*, 13(9), pp. 613–626. doi: 10.1038/nrg3207.

Sproul, D. *et al.* (2011) 'Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer', *Proceedings of the National Academy of Sciences*, 108(11), pp. 4364–4369. doi: 10.1073/pnas.1013224108.

Sproul, D. *et al.* (2012a) 'Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns', *Genome Biology*, 13(10), p. R84. doi: 10.1186/gb-2012-13-10-r84.

Sproul, D. *et al.* (2012b) 'Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns', *Genome Biology*, 13(10), p. R84. doi: 10.1186/gb-2012-13-10-r84.

Sproul, D. and Meehan, R. R. (2013) 'Genomic insights into cancer-associated aberrant CpG island hypermethylation.', *Briefings in functional genomics*. Oxford University Press, 12(3), pp. 174–90. doi: 10.1093/bfgp/els063.

Spruijt, C. G. *et al.* (2013) 'Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives', *Cell*, 152(5), pp. 1146–1159. doi: 10.1016/j.cell.2013.02.004.

Stavrou, S. and Ross, S. R. (2015) 'APOBEC3 Proteins in Viral Immunity.', *Journal of immunology (Baltimore, Md. : 1950)*. NIH Public Access, 195(10), pp. 4565–70. doi: 10.4049/jimmunol.1501504.

Strogantsev, R. *et al.* (2015) 'Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression', *Genome Biology*, 16(1), p. 112. doi: 10.1186/s13059-015-0672-7.

Sugiyama, Y. *et al.* (2010) 'The DNA-binding activity of mouse DNA methyltransferase 1 is regulated by phosphorylation with casein kinase 1δ/ε', *Biochemical Journal*, 427(3), pp. 489–497. doi: 10.1042/BJ20091856.

Sultana, R. *et al.* (2002) 'Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins.', *Genomics*, 80(2), pp. 129–34. doi: 10.1006/GENO.2002.6810.

Sun, L. *et al.* (2007) 'Phosphatidylinositol 3-kinase/protein kinase B pathway stabilizes DNA methyltransferase I protein and maintains DNA methylation', *Cellular Signalling*, 19(11), pp. 2255–2263. doi: 10.1016/j.cellsig.2007.06.014.

Sutendra, G. *et al.* (2014) 'A Nuclear Pyruvate Dehydrogenase Complex Is Important for the Generation of Acetyl-CoA and Histone Acetylation', *Cell*, 158(1), pp. 84–97. doi: 10.1016/j.cell.2014.04.046.

Suzuki, M. M. and Bird, A. (2008) 'DNA methylation landscapes: provocative insights from epigenomics', *Nature Reviews Genetics*, 9(6), pp. 465–476. doi: 10.1038/nrg2341.

Tachibana, M. *et al.* (2002) 'G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis', *Genes & Development*, 16(14), pp. 1779–1791. doi: 10.1101/gad.989402.

Tachibana, M. *et al.* (2008) 'G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription', *The EMBO Journal*, 27(20), pp. 2681–2690. doi: 10.1038/emboj.2008.192.

Tadokoro, Y. *et al.* (2007) 'De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells', *The Journal of Experimental Medicine*, 204(4), pp. 715–722. doi: 10.1084/jem.20060750.

Tahiliani, M. *et al.* (2009) 'Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1', *Science*, 324(5929), pp. 930–935. doi: 10.1126/science.1170116.

Takashima, Y. *et al.* (2014) 'Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human', *CELL*, 158, pp. 1254–1269. doi: 10.1016/j.cell.2014.08.029.

Takebayashi, S. *et al.* (2007) 'Major and essential role for the DNA methylation mark in mouse embryogenesis and stable association of DNMT1 with newly replicated regions.', *Molecular and cellular biology*. American Society for Microbiology (ASM), 27(23), pp. 8243–58. doi: 10.1128/MCB.00899-07.

Takeshita, K. *et al.* (2011) 'Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1)', *Proceedings of the National Academy of Sciences*, 108(22), pp. 9055–9059. doi: 10.1073/pnas.1019629108.

Taleahmad, S. *et al.* (2016) 'Proteome Analysis of Ground State Pluripotency', *Scientific Reports*. Nature Publishing Group, 5(1), p. 17985. doi: 10.1038/srep17985.

Tamm, C., Pijuan Galitó, S. and Annerén, C. (2013) 'A Comparative Study of Protocols for Mouse Embryonic Stem Cell Culturing', *PLoS ONE*. Edited by D. S. Milstone, 8(12), p. e81156. doi: 10.1371/journal.pone.0081156.

Thijssen, P. E. *et al.* (2015) 'Mutations in CDCA7 and HELLS cause immunodeficiency–centromeric instability–facial anomalies syndrome', *Nature Communications*, 6(1), p. 7870. doi: 10.1038/ncomms8870.

Thompson, P. J., Macfarlan, T. S. and Lorincz, M. C. (2016) 'Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire', *Molecular Cell*, 62(5), pp. 766–776. doi: 10.1016/j.molcel.2016.03.029.

Thomson, J. P. *et al.* (2010) 'CpG islands influence chromatin structure via the CpG-binding protein Cfp1', *Nature*, 464(7291), pp. 1082–1086. doi: 10.1038/nature08924.

Thorvaldsen, J. L., Duran, K. L. and Bartolomei, M. S. (1998) 'Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2', *Genes & Development*, 12(23), pp. 3693–3702. doi: 10.1101/gad.12.23.3693.

Tong, L. (2013) 'Structure and function of biotin-dependent carboxylases.', *Cellular and molecular life sciences : CMLS*. NIH Public Access, 70(5), pp. 863–91. doi: 10.1007/s00018-012-1096-0.

Torrisani, J. *et al.* (2007) 'AUF1 cell cycle variations define genomic DNA methylation by regulation of DNMT1 mRNA stability.', *Molecular and cellular biology*. American Society for Microbiology Journals, 27(1), pp. 395–410. doi: 10.1128/MCB.01236-06.

Trinkle-Mulcahy, L. (2019) 'Recent advances in proximity-based labeling methods for interactome mapping', *F1000Research*, 8, p. 135. doi: 10.12688/f1000research.16903.1.

Tsumura, A. *et al.* (2006a) 'Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b', *Genes to Cells*, 11(7), pp. 805–814. doi: 10.1111/j.1365-2443.2006.00984.x.

Tsumura, A. *et al.* (2006b) 'Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b', *Genes to Cells*, 11(7), pp. 805–814. doi: 10.1111/j.1365-2443.2006.00984.x.

Tucker, K. L. *et al.* (1996) 'Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes.', *Genes & Development*, 10(8), pp. 1008–1020. doi: 10.1101/gad.10.8.1008.

Tuorto, F. *et al.* (2012) 'RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis', *Nature Structural & Molecular Biology*, 19(9), pp. 900–905. doi: 10.1038/nsmb.2357.

Ulrey, C. L. *et al.* (2005) 'The impact of metabolism on DNA methylation', *Human Molecular Genetics*, 14(suppl_1), pp. R139–R147. doi: 10.1093/hmg/ddi100.

Valesini, G. *et al.* (2015) 'Citrullination and autoimmunity', *Autoimmunity Reviews*, 14(6), pp. 490–497. doi: 10.1016/j.autrev.2015.01.013.

Varley, K. E. *et al.* (2013) 'Dynamic DNA methylation across diverse human cell lines and tissues', *Genome Research*, 23(3), pp. 555–567. doi: 10.1101/gr.147942.112.

Varnaitė, R. and MacNeill, S. A. (2016) 'Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID.', *Proteomics*. Wiley-Blackwell, 16(19), pp. 2503–2518. doi: 10.1002/pmic.201600123.

Vilkaitis, G. *et al.* (2001) 'The mechanism of DNA cytosine-5 methylation. Kinetic and mutational dissection of HhaI methyltransferase', *Journal of Biological Chemistry*, 276(24), pp. 20924–20934. doi: 10.1074/jbc.M101429200.

von Meyenn, F. *et al.* (2016a) 'Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells', *Molecular Cell*, 62(6), pp. 848–861. doi: 10.1016/j.molcel.2016.04.025.

von Meyenn, F. *et al.* (2016b) 'Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells', *Molecular Cell*, 62(6), pp. 848–861. doi: 10.1016/j.molcel.2016.04.025.

De Vos, M. *et al.* (2014) 'Poly(ADP-ribose) Polymerase 1 (PARP1) Associates with E3 Ubiquitin-Protein Ligase UHRF1 and Modulates UHRF1 Biological Functions', *Journal of Biological Chemistry*, 289(23), pp. 16223–16238. doi: 10.1074/jbc.M113.527424.

Vossenaar, E. R. *et al.* (2003) 'PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease', *BioEssays*, 25(11), pp. 1106–1118. doi: 10.1002/bies.10357.

Wade, P. A. (2001) 'Methyl CpG binding proteins: coupling chromatin architecture to gene regulation', *Oncogene*, 20(24), pp. 3166–3173. doi: 10.1038/sj.onc.1204340.

Walsh, C. P. and Bestor, T. H. (1999) 'Cytosine methylation and mammalian development.', *Genes & development*. Cold Spring Harbor Laboratory Press, 13(1), pp. 26–34. doi: 10.1101/gad.13.1.26.

Walsh, C. P., Chaillet, J. R. and Bestor, T. H. (1998) 'Transcription of IAP endogenous retroviruses is constrained by cytosine methylation', *Nature Genetics*, 20(2), pp. 116–117. doi: 10.1038/2413.

Walter, M. *et al.* (2016a) 'An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells', *eLife*, 5. doi: 10.7554/eLife.11418.

Walter, M. *et al.* (2016b) 'An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells', *eLife*, 5. doi: 10.7554/eLife.11418.

Wang, L. *et al.* (2017) 'PADI2-Mediated Citrullination Promotes Prostate Cancer Progression', *Cancer Research*, 77(21), pp. 5755–5768. doi: 10.1158/0008-5472.CAN-17-0150.

Wang, S. and Wang, Y. (2013) 'Peptidylarginine deiminases in citrullination, gene regulation, health and pathogenesis', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(10), pp. 1126–1135. doi: 10.1016/j.bbagrm.2013.07.003.

Wang, Y. *et al.* (2009) 'Histone hypercitrullination mediates chromatin decondensation and neutrophil extracellular trap formation', *The Journal of Cell Biology*, 184(2), pp. 205–213. doi: 10.1083/jcb.200806072.

Weber, M. *et al.* (2005) 'Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells', *Nature Genetics*, 37(8), pp. 853–862. doi: 10.1038/ng1598.

Weber, M. *et al.* (2007) 'Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome', *Nature Genetics*, 39(4), pp. 457–466. doi: 10.1038/ng1990.

Webster, K. E. *et al.* (2005) 'Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis', *Proceedings of the National Academy of Sciences*, 102(11), pp. 4068–4073. doi: 10.1073/pnas.0500702102.

Wray, J. *et al.* (2011) 'Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation', *Nature Cell Biology*, 13(7), pp. 838–845. doi: 10.1038/ncb2267.

Wray, J., Kalkan, T. and Smith, A. G. (2010) 'The ground state of pluripotency', *Biochemical Society Transactions*, 38(4), pp. 1027–1032. doi: 10.1042/BST0381027.

Wu, H. *et al.* (2010) 'Structural Biology of Human H3K9 Methyltransferases', *PLoS ONE*. Edited by N. Gay, 5(1), p. e8570. doi: 10.1371/journal.pone.0008570.

Wu, X. and Zhang, Y. (2017) 'TET-mediated active DNA demethylation: mechanism, function and beyond', *Nature Reviews Genetics*, 18(9), pp. 517–534. doi: 10.1038/nrg.2017.33.

Xu, G.-L. *et al.* (1999) 'Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene', *Nature*, 402(6758), pp. 187–191. doi: 10.1038/46052.

Xu, J. *et al.* (2009) 'Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells', *Genes & Development*, 23(24), pp. 2824–2838. doi: 10.1101/gad.1861209.

Xu, Y. *et al.* (2016) 'Mutations in PADI6 Cause Female Infertility Characterized by Early Embryonic Arrest', *The American Journal of Human Genetics*, 99(3), pp. 744–752. doi: 10.1016/j.ajhg.2016.06.024.

Yamaguchi, L. *et al.* (2017) 'Usp7-dependent histone H3 deubiquitylation regulates maintenance of DNA methylation', *Scientific Reports*, 7(1), p. 55. doi: 10.1038/s41598-017-

00136-5.

Yang, J. *et al.* (2017) 'PIM1 induces cellular senescence through phosphorylation of UHRF1 at Ser311', *Oncogene*, 36(34), pp. 4828–4842. doi: 10.1038/onc.2017.96.

Yang, N. and Xu, R.-M. (2013) 'Structure and function of the BAH domain in chromatin biology', *Critical Reviews in Biochemistry and Molecular Biology*, 48(3), pp. 211–221. doi: 10.3109/10409238.2012.742035.

Yarychkivska, O., Shahabuddin, Z., *et al.* (2018) 'BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate de novo and maintenance methylation in vivo', *Journal of Biological Chemistry*, 293(50), pp. 19466–19475. doi: 10.1074/jbc.RA118.004612.

Yarychkivska, O., Tavana, O., *et al.* (2018) 'Independent functions of DNMT1 and USP7 at replication foci', *Epigenetics & Chromatin*, 11(1), p. 9. doi: 10.1186/s13072-018-0179-z.

Ye, F. *et al.* (2018) 'Biochemical Studies and Molecular Dynamic Simulations Reveal the Molecular Basis of Conformational Changes in DNA Methyltransferase-1', *ACS Chemical Biology*, 13(3), pp. 772–781. doi: 10.1021/acschembio.7b00890.

Yin, Y. *et al.* (2017) 'Impact of cytosine methylation on DNA binding specificities of human transcription factors', *Science*, 356(6337), p. eaaj2239. doi: 10.1126/science.aaj2239.

Ying, Q.-L. *et al.* (2008a) 'The ground state of embryonic stem cell self-renewal', *Nature*, 453(7194), pp. 519–523. doi: 10.1038/nature06968.

Ying, Q.-L. *et al.* (2008b) 'The ground state of embryonic stem cell self-renewal', *Nature*, 453(7194), pp. 519–523. doi: 10.1038/nature06968.

Ying, Q. L. *et al.* (2003) 'BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3.', *Cell*, 115(3), pp. 281–92. doi: 10.1016/s0092-8674(03)00847-x.

Yoder, J. A., Walsh, C. P. and Bestor, T. H. (1997) 'Cytosine methylation and the ecology of intragenomic parasites.', *Trends in genetics : TIG*, 13(8), pp. 335–40.

Yu, D.-H. *et al.* (2013) 'Developmentally Programmed 3' CpG Island Methylation Confers Tissue- and Cell-Type-Specific Transcriptional Activation', *Molecular and Cellular Biology*, 33(9), pp. 1845–1858. doi: 10.1128/MCB.01124-12.

Yurttas, P. *et al.* (2008) 'Role for PADI6 and the cytoplasmic lattices in ribosomal storage in oocytes and translational control in the early mouse embryo', *Development*, 135(15), pp. 2627–2636. doi: 10.1242/dev.016329.

Zemach, A. *et al.* (2010) 'Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation', *Science*, 328(5980), pp. 916–919. doi: 10.1126/science.1186366.

Zhang, G. *et al.* (2015) 'Small RNA-mediated DNA (cytosine-5) methyltransferase 1 inhibition leads to aberrant DNA methylation', *Nucleic Acids Research*, 43(12), pp. 6112–6124. doi: 10.1093/nar/gkv518.

Zhang, H. *et al.* (2016) 'A cell cycle-dependent BRCA1–UHRF1 cascade regulates DNA double-strand break repair pathway choice', *Nature Communications*. Nature Publishing Group, 7(1), p. 10201. doi: 10.1038/ncomms10201.

Zhang, T. *et al.* (2016) 'G9a/GLP Complex Maintains Imprinted DNA Methylation in

Embryonic Stem Cells', *Cell Reports*, 15(1), pp. 77–85. doi: 10.1016/j.celrep.2016.03.007.

Zhang, X. *et al.* (2011) 'Genome-Wide Analysis Reveals PADI4 Cooperates with Elk-1 to Activate c-Fos Expression in Breast Cancer Cells', *PLoS Genetics*. Edited by A. Akhtar. Public Library of Science, 7(6), p. e1002112. doi: 10.1371/journal.pgen.1002112.

Zhang, X. *et al.* (2012) 'Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination facilitates estrogen receptor target gene activation', *Proceedings of the National Academy of Sciences*, 109(33), pp. 13331–13336. doi: 10.1073/pnas.1203280109.

Zhang, X. *et al.* (2016) 'Peptidylarginine deiminase 1-catalyzed histone citrullination is essential for early embryo development', *Scientific Reports*, 6(1), p. 38727. doi: 10.1038/srep38727.

Zhang, Y. *et al.* (2019) 'TurboID-based proximity labeling reveals that UBR7 is a regulator of N NLR immune receptor-mediated immunity', *Nature Communications*. Nature Publishing Group, 10(1), p. 3252. doi: 10.1038/s41467-019-11202-z.

Zhang, Yunfang *et al.* (2018) 'Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs', *Nature Cell Biology*, 20(5), pp. 535–540. doi: 10.1038/s41556-018-0087-2.

Zhang, Z.-M. *et al.* (2015) 'Crystal Structure of Human DNA Methyltransferase 1', *Journal of Molecular Biology*, 427, pp. 2520–2531. doi: 10.1016/j.jmb.2015.06.001.

Zhao, Q. *et al.* (2016a) 'Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals', *Nature Communications*, 7(1), p. 12464. doi: 10.1038/ncomms12464.

Zhao, Q. *et al.* (2016b) 'Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals', *Nature Communications*, 7(1), p. 12464. doi: 10.1038/ncomms12464.

Ziller, M. J. *et al.* (2018) 'Dissecting the Functional Consequences of De Novo DNA Methylation Dynamics in Human Motor Neuron Differentiation and Physiology', *Cell Stem Cell*, 22(4), pp. 559-574.e9. doi: 10.1016/j.stem.2018.02.012.

Zylicz, J. J. *et al.* (2015) 'Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development', *eLife*, 4. doi: 10.7554/eLife.09571.